# استخراج خط سیر حرکتی وسایل نقلیه از طریق تشخیص و ردیابی بلادرنگ اشیا در سامانه‌های نظارتی مبتنی بر تحلیل ویدئویی

**مقاله علمی-پژوهشی**

*میلاد بیات (نویسنده مسئول)، دانش آموخته کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران

*پست الکترونیکی نویسنده مسئول: MiladBayatAcademic@gmail.com

دریافت: ۱۴۰۵/۰۳/۰۷- پذیرش: ۱۴۰۵/۰۸/۰۵

صفحه ۱۱۸-۹۷

**چکیده**

*بخشی از جاده‌های کشور دارای دوربین‌های نظارتی هستند. این دوربین‌ها توسط اپراتورهای انسانی به طور پیوسته پایش می‌شوند. باتوجه‌به تعداد زیاد دوربین‌ها، این روش پایش ترافیک علاوه بر طاقت‌فرسا بودن و هزینه زیاد، دارای خطای بالایی است. بنابراین، نیاز به یک چارچوب خودکار با قابلیت پردازش بلادرنگ برای استخراج خط سیر حرکتی وسایل نقلیه وجود دارد که پژوهشگران و متخصصان حمل‌ونقل بتوانند از آن برای بررسی رفتارهای خردنگر وسایل نقلیه در زمینه‌های مختلف استفاده کنند. هدف از این پژوهش، طراحی چارچوبی پیشرفته و کارآمد برای استخراج خط سیر حرکتی دقیق و پیوسته وسایل نقلیه از طریق تشخیص و ردیابی بلادرنگ ا شیا در سامانه‌های نظارتی مبتنی بر تحلیل ویدئویی ا ست. در این مطالعه، از الگوریتم YOLOv10 برای ت شخیص ا شیا و ByteTrack برای ردیابی آن‌ها بهره گرفته شده است. با درک اهمیت دسترسی به مجموعه‌داده‌ای متنوع، باکیفیت و نماینده، مجموعه‌داده‌ای تحت عنوان Surveillance Cameras با هدف برچسب‌گذاری باکیفیت از صحنه‌های مختلف ترافیکی، تهیه گردید. از تکنیک یادگیری انتقالی برای تنظیم دقیق مدل از پیش آموزش‌دیده YOLOv10m بر روی این مجموعه‌داده بهره گرفته شده است. مدل تنظیم دقیق شده توانسته در معیارهای دقت، بازیابی و میانگین دقت متوسط به ترتیب به مقادیر ۹۶,۳، ۹۵,۲ و ۹۸,۳ درصد دست یابد و با تأخیر ۵,۰۸ میلی‌ثانیه اجرا شود. نتایج حاصل نشان می‌دهد که چارچوب پیشنهادی قادر است اکثر وسایل نقلیه را به طور دقیق شناسایی و به‌درستی دسته‌بندی کرده و نرخ فریم بالایی را در کاربردهای بلادرنگ تضمین کند. این عملکرد، در رویکرد ردیابی مبتنی بر تشخیص، فرایند تطبیق داده را بهبود بخشیده و ضمن حفظ پیوستگی شناسه‌ها در فریم‌های متوالی، امکان استخراج خط سیر حرکتی دقیق، پایدار و پیوسته را فراهم می‌سازد. چارچوب پیشنهادی اطلاعات دقیقی درباره تغییر مکان هر شناسه در طول زمان، ارائه داده و در تحلیل الگوهای حرکتی، پیش‌بینی مکان، استخراج پارامترهای ترافیکی و شناسایی تداخلات احتمالی کاربرد دارد.*

**واژه‌های کلیدی:** استخراج خط سیر حرکتی، بینایی کامپیوتری، پردازش بلادرنگ، ردیابی مبتنی بر تشخیص، یادگیری عمیق

## ۱-مقدمه

روش جمع‌آوری داده‌ها برای تحلیل یک عامل کلیدی است که زمان، هزینه و کارایی نتایج را تعیین می‌کند (Azimjonov & Özmen, 2021). داده‌ها از طریق مجموعه‌ای از حسگرها جمع‌آوری شده که هر کدام مزایای منحصربه‌فردی دارند و به شرایط مختلف سازگارند. دوربین یکی از رایج‌ترین روش‌ها برای نظارت بر ترافیک است، زیرا داده‌های بصری مشابه مشاهدات انسانی را فراهم می سازد (H. Yang, 2023). در حال حا ضر بخ شی از جاده‌های کشور به دوربین‌های نظارتی مجهز هستند که به‌صورت راهبردی در نقاط کلیدی نظیر تقاطع‌ها نصب

شـــده و هدف آن‌ها کمک به نهادهای مدیریت ترافیک از طریق ارائه اطلاعات لحظه‌ای اســـت (T. Li et al., 2024). این دوربین‌ها توســـط اپراتورهای انســـانی به طور پیو سته پایش می شوند. باتوجه‌به تعداد زیاد دوربین‌ها، این روش پایش ترافیک علاوه بر طاقت‌فرسا بودن و هزینه زیاد، دارای خطای بالایی است. در مقابل، استفاده از رادار، لایدر، امواج رادیویی یا لیزر برای حل این مشـــکل، می‌توا ند هزینه‌بر، زمان‌بر و دشـــوار باشـــد (Azimjonov & Özmen, 2021). چالش اصلی دوربین‌ها تأثیرپذیری زیاد از شـــرایط محیطی، ازجمله وضـــعیت آب‌وهوا و کاهش عملکرد در شرایط نور کم است (Bunaiya Prayudha, 2021; Hassaballah et al., 2021). جـمـع‌آوری داده‌های بصـــری به‌تنهایی کافی نیســـت، این داده‌ها باید پردازش شوند تا اطلاعات مفیدی برای استفاده در مطالعات حمـل‌ونقـل اســـتخراج شـــود (Emiyah, 2021). در سامانه‌های خودکار نظارت ترافیک و سامانه‌های حمل‌ونقل هوشمند، بینایی کامپیوتری قادر است به طور هم‌زمان طیف وسیعی از اطلاعات را از صحنه‌های ترافیکی استخراج کند (Rezaei et al., 2023; Zhou et al., 2021). بسیاری از کاربردهای بینایی کامپیوتری نیازمند تشـــخیص و ردیابی فوری اشیا و با سرعتی برابر یا بیشتر از نرخ ورود داده‌های جدید هســـتند. این ســـامانه‌ها به‌عنوان بلادرنگ شـــناخته می‌شـــوند (Christián, 2023; Mattingly, 2023; Wang et al., 2024). با پردازش بلادرنگ، الگوریتم‌های بینایی کامپیوتری می‌توانند وســـایل نقلیه را به‌ســـرعت شنا سایی و د سته‌بندی کنند و امکان تصمیمات و اقدامات به‌موقع را فراهم ســـازند (Christián, 2023). بنابراین، نیاز به یک چارچوب خودکار با قابلیت پردازش بلادرنگ برای استخراج خط سیر حرکتی وسایل نقلیه وجود دارد که پژوهشـــگران و متخصـــصـــان حمل‌ونقل بتوانند از آن برای بررسی رفتارهای خردنگر وسایل نقلیه در زمینه‌های مختلف استفاده کنند. در این راستا، توانایی تشخیص و ردیابی اشیا یکی از نیازهای اساسی هر سامانه نظارت بر ترافیک است که امکان دستیابی به قابلیت‌های پیشرفته‌تر مانند استخراج خط سیر حرکتی وسـایل نقلیه را فراهم می‌آورد (Ding, 2023).

بســـیاری از چارچوب‌های تحلیل صـــحنه از رویکردهای «تشخیص–دسته‌بندی–ردیابی» یا «تشخیص–خوشه‌بندی–ردیابی» پیروی می‌کنند. از این رو، الگوریتم تشـــخیصـــی که توســـط یک چارچوب مشـــخص به کار گرفته می‌شـــود، می‌تواند تأثیر قابل‌توجهی بر کیفیت تحلیل و تعداد اشـــیا نادرست یا ازدست‌رفته داشته باشد (Hislop-Lynch & Richard, 2023). الگوریتم‌های تشخیص اشیا مبتنی بر یادگیری عمیق، در اســـتخراج اطلاعات وســـایل نقلیه مانند ظاهر، رنگ، نوع و مکان از طریق دوربین‌های نظارتی موفق بوده‌انـد (Hassaballah et al., 2021; H. Yang, 2023). بااین‌حال، هزینه‌های ذخیره‌سازی داده، ارتباطات و محاســـبات، اجرای عملی این ســـامانه‌ها را محدود می‌کند (H. Yang, 2023). همچنین، عملکرد آن‌ها به شدت در صحنه‌های ترافیکی با انسداد زیاد (Abdel-Aty et al., 2023; Ngeni et al., 2024; Rezaei et al., 2023)، تغییرات در ظاهر وسـایل نقلیه (Muhib et al., 2023; Ngeni et al., 2024; H. Yang, 2023)، تغییرات ناگهانی محیط (Abdel-Aty et al., 2023; Rezaei et al., 2023) و شرایط نوری نامناسب (Abdel-Aty et al., 2023; Bakirci, 2024; Rezaei et al., 2023) محـدود می‌شـــود. از جملـه چـالش‌هـایی کـه الگوریتم‌های ردیابی اشیا با آن مواجه‌اند، می‌توان به انسداد جزئی یا کامل (Derek Adam Gloudemans et al., 2023; Muhib et al., 2023)، حـرکـت ســـریـع (Messoussi et al., 2021)، تغییرات مقیاس (Abdel-Aty et al., 2023; Messoussi et al., 2021)، تغییرات شرایط نوری، شباهت اهداف (Messoussi et al., 2021)، تـراکـم زیـاد (O. Boadu, 2024; Siddique, 2023; Sousa, 2023) و ناهنجاری‌های رفتاری (Sousa, 2023) اشـــاره کرد که این عوامل می‌توانند به ازدســـت‌رفتن تشـــخیص‌ها، تداخل مســـیرهای ردیابی، تغییر شناسه، ایجاد شکاف در مسیرهای ردیابی و کـاهش عملکرد منجر شـــونـد (Messoussi et al., 2021). چالش دیگر، تشـــخیص و ردیابی بلادرنگ اشـــیا اســـت که نیاز به الگوریتم‌های ســـریع و کارآمد دارد (Christián, 2023).

زیرا روش‌های متعارف محاسبات سنگینی دارند و در بیشتر موارد قادر به بهره‌گیری از اطلاعات پیشین اشیا یا الگوهای موجود در صحنه برای افزایش سرعت تشخیص و ردیابی نیستند (Derek Adam Gloudemans et al., 2023). بنابراین، بهره‌گیری از الگوریتم های تشخیص و رد یابی بلادرنگ و پایدار به‌منظور توسعه ابزارهای تحلیلی مؤثر و کارآمد، امری ضروری است. برای آموزش و ارزیابی الگوریتم‌های تشخیص و ردیابی اشیا و غلبه بر برخی از این مشکلات، دسترسی به یک مجموعه‌داده متنوع، باکیفیت و نماینده برای اطمینان از توانایی تعمیم مدل در صحنه‌های مختلف، حیاتی است (Bakirci, 2024; O. Boadu, 2024).

هدف از این پژوهش، طراحی چارچوبی پیشرفته و کارآمد برای استخراج خط سیر حرکتی دقیق و پیوسته وسایل نقلیه از طریق تشخیص و ردیابی بلادرنگ اشیا در سامانه‌های نظارتی مبتنی بر تحلیل ویدئویی است. با درک اهمیت دسترسی به مجموعه‌داده‌ای متنوع، باکیفیت و نماینده، مجموعه‌داده‌ای جدید تحت عنوان Surveillance Cameras با هدف برچسب‌گذاری باکیفیت از صحنه‌های مختلف ترافیکی، تهیه گردید. ابتدا، از تکنیک یادگیری انتقالی برای تنظیم دقیق مدل از پیش آموزش‌دیده YOLOv10m بر روی این مجموعه‌داده بهره گرفته شده و مدل تنظیم دقیق شده YOLOv10m برای شناسایی و استخراج ویژگی‌های وسایل نقلیه در هر فریم به کار گرفته می‌شود. سپس، با استفاده از الگوریتم ردیابی ByteTrack، تطبیق میان فریم‌های متوالی برقرار می‌شود. این الگوریتم با تخصیص شناسه‌های منحصربه‌فرد به هر وسیله نقلیه و ذخیره موقعیت آن در فریم های مختلف، امکان رد یابی پیوسته و دقیق را فراهم می‌آورد. درنهایت، با استفاده از داده‌های مکانی به‌دست‌آمده، خط سیر حرکتی هر وسیله نقلیه استخراج می‌شود.

## ۲-پیشینه تحقیق

بخش دوم این مقاله به بررسی مبانی نظری و پیشینه تحقیقاتی مرتبط با موضوع تحقیق پرداخته است. در این بخش، ابتدا مفاهیم بنیادی بینایی کامپیوتری و یادگیری عمیق مورد بررسی قرار می‌گیرند. سپس، به تکنیک‌های تشخیص و ردیابی اشیا با تمرکز بر یادگیری عمیق پرداخته می‌شود. در نهایت، تحقیقات و کارهای مشابه در این حوزه تحلیل می‌شود تا مبنای علمی و نظری لازم برای ادامه تحقیق حاضر فراهم گردد.

### ۲-۱-بینایی کامپیوتری

شبکه‌های عصبی مصنوعی توانایی مغز انسان را در پردازش و تفسیر داده‌های بصری شبیه‌سازی می‌کنند (Muhib et al., 2023; O. Boadu, 2024). به‌عنوان یک زیرمجموعه از هوش مصنوعی، بینایی کامپیوتری توانایی ماشین‌ها را برای رمزگشایی و اقدام بر اساس اطلاعات استخراج‌شده از منابع بصری تسهیل می‌کند (Christián, 2023). پیشرفت های اخیر در قدرت پردازش، ظرفیت حافظه و رزولوشن حسگرهای تصویر به بهبود عملکرد و پیچیدگی زمانی الگوریتم‌های بینایی کامپیوتری کمک کرده است (Patil, 2021). روش‌های مدرن بینایی کامپیوتری عمدتاً با بهره‌گیری از یادگیری ماشین به‌عنوان یک مؤلفه اصلی، به حل مسائل دنیای واقعی می‌پردازند.

مدل‌های یادگیری ماشین برای وظایف بینایی کامپیوتری با استفاده از تکنیک‌های متنوعی از جمله یادگیری با نظارت، یادگیری خودنظارتی، یادگیری تقویتی، یادگیری نیمه‌نظارتی و یادگیری بدون نظارت آموزش داده می‌شوند (Siddique, 2023).

### ۲-۲-یادگیری عمیق

یادگیری عمیق زیرمجموعه‌ای از یادگیری ماشین است که از ساختار چندلایه‌ای شبکه‌های عصبی تشکیل شده و سعی دارد سلسله‌مراتبی از ویژگی‌ها را از داده‌های ورودی بیاموزد (Ait Ouallane et al., 2022; Sousa, 2023). واژه "عمیق" به ساختار لایه‌ای این شبکه‌ها اشاره دارد (Stryk, 2023). یادگیری عمیق به‌واسطه توانایی پردازش حجم عظیمی از داده‌ها، نیاز کمتر به مداخله انسان در مهندسی داده، توانایی یادگیری الگو های پیچیده و عملکرد بهتر نسبت به شبکه‌های عصبی سطحی، توجه زیادی را به خود جلب کرده است (Heidari et al., 2022). بسیاری از کاربردها و خدمات هوش مصنوعی تو سط یادگیری عمیق پشتیبانی می‌شوند که با انجام وظایف شناختی و فیزیکی

بدون نیاز به مداخله انسانی، فرایند اتوماسیون را تقویت می‌کند.

### ۲-۳-شبکه عصبی پیچشی

شبکه‌های عصبی پیچشی با استخراج ویژگی‌های مکانی و محلی از نواحی مجاور، الگوهای موجود در تصاویر را یاد گرفته (Christián, 2023) و حداقل در یکی از لایه‌های خود، به‌جای ضرب ماتریسی از عملیات کانولوشن استفاده می‌کنند. کانولوشن یک عملیات ریاضی بر روی دو تابع است که تأثیر یک تابع بر تابع دیگر را بیان می‌کند. این عملیات را می‌توان به‌صورت معادله (۱) نوشت که در آن $f(t)$ ورودی، $g(t)$ فیلتر و خروجی $s(t)$ به‌عنوان نقشه ویژگی شناخته می‌شود. در صورت استفاده از داده‌های گسسته، انتگرال را می‌توان با سیگما جایگزین کرد (معادله (۲)) (Stryk, 2023).

$$s(t) = f(t) \times g(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \quad (۱)$$

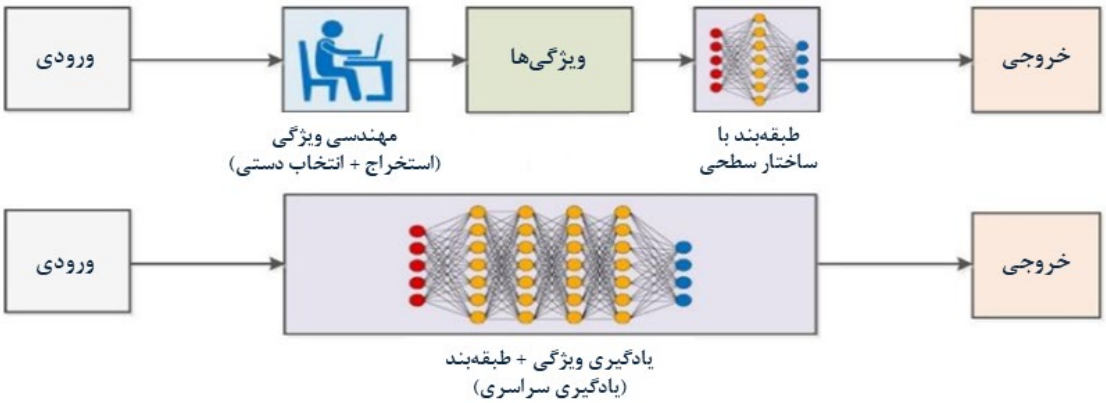
$$s(t) = f(t) \times g(t) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(t-\tau) \quad (۲)$$

این نوع شبکه برای پردازش داده‌هایی با ساختار شبکه‌ای، مانند تصاویر (Christián, 2023; Siddique, 2023; Stryk, 2023) یا حتی داده‌های سری زمانی که بتوان آن‌ها را به‌عنوان یک شبکه $N \times ۱$ در نظر گرفت، استفاده می‌شود (Stryk, 2023). در طول آموزش، شبکه‌های عصبی پیچشی پارامترهای خود (وزن‌ها و بایاس) را برای کاهش اختلاف بین خروجی‌های پیش‌بینی شده و برچسب‌های واقعی تنظیم می‌کنند. این فرایند بهینه‌سازی، به طور مکرر پارامترها را به‌روزرسانی می‌کند تا عملکرد شبکه بهبود یابد (Christián, 2023).

### ۲-٤-یادگیری انتقالی

یادگیری انتقالی روشی برای ایجاد آموزش‌های اختصاصی جدید و تسریع فرایند با استفاده از اطلاعات آموزش پیشین است. تان و همکاران یادگیری انتقالی را به‌عنوان فرایند انتقال دانش از دامنه مبدأ به دامنه هدف توصیف می‌کنند. در این سناریو، دامنه به‌صورت $\mathcal{D} = \{\mathcal{X}, \mathrm{P(X)}\}$ تعریف می‌شود که در آن $\mathcal{X}$ فضای ویژگی و $\mathrm{P(X)}$ توزیع احتمال حاشیه‌ای است، به‌طوری که $\mathrm{X} = \{x_1, x_2, \dots, x_n\}, x_i \in \mathcal{X}$. وظیفه یادگیری به صورت $\mathcal{T} = \{\mathcal{Y}, f(x)\}$ تعریف می شود، که در آن $\mathcal{Y}$ فضای برچسب‌ها و $f(x)$ تابع پیش‌بینی هدف است. بر اساس این سناریو، با درنظرگرفتن وظیفه هدف $\mathcal{T}_\mathrm{T}$ در دامنه هدف $\mathcal{D}_\mathrm{T}$، می‌توان از دانش حاصل از دامنه مبدأ $\mathcal{D}_\mathrm{S}$ و وظیفه مبدأ $\mathcal{T}_\mathrm{S}$ برای بهبود فرایند یادگیری در وظیفه هدف استفاده کرد. هدف یادگیری انتقالی، بهبود عملکرد تابع پیش‌بینی $f_T(\cdot)$ برای وظیفه یادگیری $\mathcal{T}_\mathrm{T}$ با استفاده از دانش حاصل از $\mathcal{D}_\mathrm{S}$ و $\mathcal{T}_\mathrm{S}$ است. در این شرایط، معمولاً $\mathcal{D}_\mathrm{S} \neq \mathcal{D}_\mathrm{T}$ و/یا $\mathcal{T}_\mathrm{S} \neq \mathcal{T}_\mathrm{T}$ و همچنین در اکثر موارد تعداد نمونه‌های موجود در دامنه مبدأ $\mathcal{D}_\mathrm{S}$ بسیار بیشتر از دامنه هدف $\mathcal{D}_\mathrm{T}$ است، یعنی $N_S \gg N_T$. تان و همکاران همچنین چهار دسته از رویکردهای یادگیری انتقالی را معرفی می‌کنند که شامل روش‌های مبتنی بر نمونه، نگاشت، شبکه و یادگیری تخاصمی هستند (Bunaiya Prayudha, 2021).

در این پژوهش، از رویکرد مبتنی بر شبکه برای آموزش مدل تشخیص اشیا بهره گرفته شده است. یادگیری انتقالی مبتنی بر شبکه، نوعی یادگیری انتقالی است که از یک شبکه از پیش آموزش‌دیده در دامنه مبدأ بهره گرفته و به استفاده از معماری شبکه عصبی و بازطراحی خروجی آن برای نیازهای خاص اشاره دارد. سه راهبرد مختلف برای اجرای یادگیری انتقالی مبتنی بر شبکه وجود دارد. در اولین راهبرد، مجموعه‌داده برای آموزش کل شبکه استفاده می‌شود. ویژگی اصلی این روش این است که به یک مجموعه‌داده بزرگ و منابع محاسباتی زیاد نیاز دارد. راهبرد دوم شامل آموزش جزئی معماری شبکه است. در این روش، باید به طور مشخص تعیین شود که کدام بخش از شبکه باید آموزش داده شود و کدام بخش ثابت بماند. آخرین راهبرد تنها شامل آموزش یا تنظیم دقیق لایه‌های دسته‌بندی است، درحالی‌که لایه‌های استخراج ویژگی صرفاً برای تأمین ورودی لایه‌های دسته‌بندی مورد استفاده قرار می‌گیرند (Bunaiya Prayudha, 2021). در این راهبرد، هدف اصلاح لایه‌های دسته‌بندی، از جمله تعیین دسته اشیا و پارامترهای آموزشی است. این راهبرد زمانی مفید است که داده‌های محدودی برای وظیفه جدید در دسترس باشد و یا




**شکل ۱. روند اجرایی الگوریتم‌های سنتی در مقابل یادگیری عمیق**

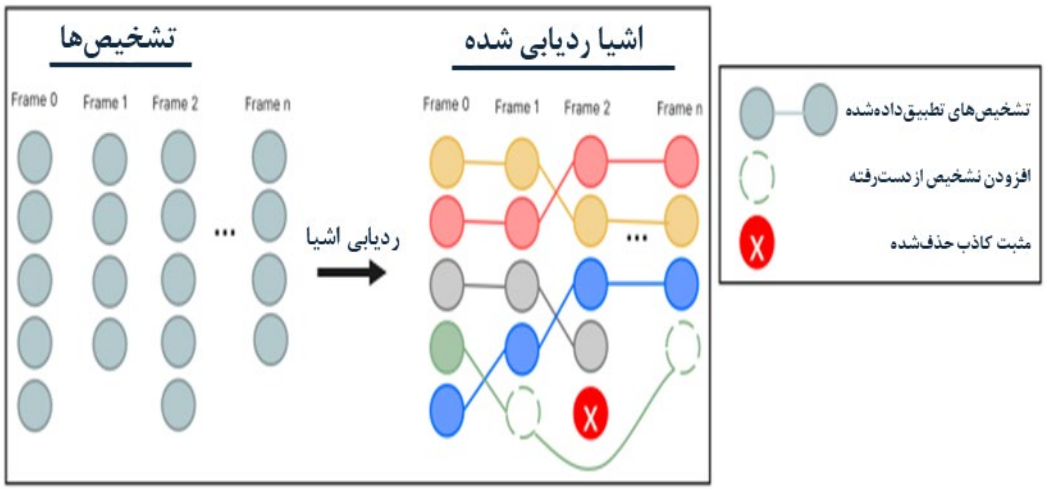
منابع محاسباتی لازم برای آموزش یک مدل از ابتدا در دسترس نباشد. علاوه‌براین، شبکه از پیش آموزش‌دیده وظیفه‌ای مشابه با وظیفه هدف داشته باشد. به طور خلاصه، یادگیری انتقالی به مدل‌ها این امکان را می‌دهد که از دانش به‌دست‌آمده از وظایف قبلی برای بهبود عملکرد در وظایف جدید بهره ببرند که این امر یادگیری ماشین را دسترس‌پذیرتر و کارآمدتر می‌کند (Bunaiya Prayudha, 2021; Muhib et al., 2023).

### ۲-۵-تشخیص اشیا

تشخیص اشیا به وظیفه شناسایی و مکان‌یابی اشیا خاص در یک تصویر یا توالی تصاویر اشاره دارد (Abdel-Aty et al., 2023; Azimjonov & Özmen, 2021; Gomaa et al., 2022). از جمله روش‌های تشخیص اشیا می‌توان به روش‌های مبتنی بر ویژگی، حرکت، ظاهر و یادگیری عمیق اشاره کرد. روش‌های مبتنی بر ویژگی به استخراج ویژگی‌های دستی وابسته هستند (Abdel-Aty et al., 2023; Azimjonov & Özmen, 2021). روش‌های مبتنی بر حرکت بر روی نواحی یا پیکسل‌های متحرک تمرکز دارند (Azimjonov & Özmen, 2021) درحالی که روش‌های مبتنی بر ظاهر ویژگی‌هایی نظیر رنگ، بافت و شکل را استخراج کرده (J. Li et al., 2021) و با استفاده از الگوهای مرجع یا طبقه‌بندهای سنتی به تشخیص اشیا می‌پردازند (Christián, 2023). در الگوریتم‌های سنتی تشخیص اشیا معمولاً استخراج و انتخاب ویژگی‌ها به‌صورت دستی انجام می‌شود (Mehrizi et al., 2021). در حالی که یادگیری عمیق، فرایند یادگیری سراسری را معرفی می‌کند که در آن شبکه‌های عصبی عمیق الگوهای نهفته در تصاویر آموزشی را کشف کرده و ویژگی‌های توصیفی خاص هر دسته از اشیا را شناسایی می‌کنند. روند اجرایی معماری الگوریتم‌های سنتی و یادگیری عمیق در شکل ۱ نشان داده شده است (Patil, 2021). الگوریتم‌های تشخیص اشیا مبتنی بر یادگیری عمیق مدت‌هاست که به‌عنوان روش ترجیحی با استفاده از شبکه‌های عصبی پیچشی و داده‌های برچسب‌گذاری شده مورد استفاده قرار می‌گیرند و به دو دسته روش‌های یک مرحله‌ای و دو مرحله‌ای تقسیم می‌شوند (Hassaballah et al., 2021; J. Li et al., 2021; Rezaei et al., 2023).

#### ۲-۵-۱-شناساگر یک مرحله‌ای

شناساگرهای یک مرحله‌ای کادرهای محاطی اشیا و احتمالات دسته‌ای مرتبط با آن را مستقیماً از یک تصویر کامل در یک محاسبه واحد از طریق یک شبکه عصبی یکپارچه پیش‌بینی می‌کنند (Hnewa, 2023; Mattingly, 2023). این روش، وظایف دسته‌بندی و مکان‌یابی اشیا را در یک فرایند یکپارچه ترکیب کرده و با حذف مرحله تولید و تحلیل نواحی پیشنهادی، می‌تواند مزایای سرعتی قابل‌توجهی ارائه دهد (Abdel-Aty et al., 2023; Hassaballah et al., 2021). این امر باعث می‌شود که این مدل‌ها در مقایسه با مدل‌های دو مرحله‌ای بسیار سریع‌تر باشند و منابع کمتری مصرف کنند که این رویکرد به‌ویژه در شرایطی که پردازش بلادرنگ ضروری است، سودمند است و امکان استقرار آن‌ها را روی




**شکل ۲. مسئله ردیابی اشیا. شناسه‌های منحصربه‌فرد (رنگ‌ها) به تشخیص‌ها (دایره‌ها) اختصاص داده می‌شوند. تشخیص‌های ازدست‌رفته (دایره‌های خط‌چین) باید بازیابی شوند و تشخیص‌های مثبت کاذب (X قرمز) باید حذف شوند.**

دستگاه‌های لبه فراهم می‌آورد. ازآنجاکه این مدل‌ها در مقایسه با مدل‌های دو مرحله‌ای پیچیدگی کمتری دارند، دقت کمتری نیز دارند (Abdel-Aty et al., 2023; H. (Frank) Yang et al., 2023).

**۲-۵-۲-شناساگر دو مرحله‌ای**

در شناساگرهای دومرحله‌ای، دو وظیفهٔ اصلی دنبال می‌شود: الف) پیشنهاد ناحیه ب) دسته‌بندی و تعیین مکان (Abdel-Aty et al., 2023; Hassaballah et al., 2021). در مرحله اول یک مجموعه از نواحی کاندیدای اشیا یا نواحی موردنظر تولید می شود که نشان‌دهنده بخش‌هایی از تصویر با احتمال بالای وجود اشیا است (García-Aguilar et al., 2024; O. Boadu, 2024). سپس در مرحله دوم معماری، ویژگی‌های بریده شده برای دسته‌بندی اشیا و تنظیم دقیق کادرهای محاطی استفاده می‌شوند (Abdel-Aty et al., 2023; Mattingly, 2023; Stryk, 2023). این فرایند دو مرحله‌ای امکان تحلیل دقیق نواحی پیشنهادی را فراهم کرده و می‌تواند دقت تشخیص اشیا را به‌ویژه در سناریوهای پیچیده شامل اشیای کوچک، متعدد یا دارای تنوع بصری بالا بهبود بخشد (O. Boadu, 2024). بااین حال، این مزیت به بهای کاهش سرعت استنتاج مدل حاصل می‌شود (Abdel-Aty et al., 2023; García-Aguilar et al., 2024; Hassaballah et al., 2021; H. (Frank) Yang et al., 2023). انتخاب بین روش های یک مرحله‌ای و دو مرحله‌ای به نیازها و محدودیت‌های کاربرد موردنظر بستگی دارد و عواملی نظیر پیچیدگی صحنه، تراکم و تنوع اشیا، الزامات سرعت پردازش و محدودیت‌های منابع محا سباتی را شامل می‌شود (O. Boadu, 2024).

**۲-۶-ردیابی اشیا**

ردیابی اشیا دو هدف دارد: الف) پیش‌بینی مکان اشیا در ویدئوها ب) تطبیق اشیا شناسایی شده قبلی و مسیر ردیابی آن‌ها در فریم فعلی (شکل ۲) (Mehrizi et al., 2021; VAŇO, 2024). در اکثر مدل های رد یابی اشیا از روش‌های ردیابی مبتنی بر تشخیص و ردیابی بدون نیاز به تشخیص برای آغاز فرایند ردیابی استفاده می‌شود (Ngeni et al., 2024). در رد یابی مبتنی بر تشخیص، از یک شنا ساگر برای تشخیص ا شیا در هر فریم ا ستفاده شده و اطلاعات تشخیص به الگوریتم ردیابی ارسال می‌شود تا تطبیق اشیا از یک فریم به فریم دیگر برقرار گردد (Psalta et al., 2024; Stryk, 2023). در مقابل، ردیابی بدون نیاز به تشخیص، از مدلسازی پس‌زمینه برای تشخیص اشیا متحرک در فریم های و یدئو پیش از شروع رد یابی بهره می‌برد. بر این ا ساس، ردیابی اشیا ساده‌سازی شده و به دو د سته تقسیم می شود: ردیابی آنلاین و آفلاین (Ngeni et al., 2024). برخلاف روش‌های ردیابی آفلاین که قبل از تولید نتایج، تعداد زیادی فریم یا یک دنباله ویدئویی کامل را پردازش می‌کنند، روش های رد یابی آنلاین هر بار یک فریم را پردازش می‌کنند بدون اینکه دسترسی یا اطلاعاتی در مورد فریم‌های آینده داشته باشند.

بخش قابل‌توجهی از الگوریتم‌های ردیابی اشیا به روش‌های تطبیق داده نظیر بهینه‌سازی جریان شبکه یا الگوریتم مجارستانی وابسته‌اند. این فرایند تطبیق شامل مقایسه تشخیص‌های فریم جاری با مسیرهای فعال یا وضعیت پیش‌بینی‌شده آن‌ها برای همان فریم است. منابع اصلی اطلاعاتی که معمولاً برای تعیین هزینهٔ تطابق یا عدم تشابه در فرایند تطبیق داده مورداستفاده قرار می‌گیرند، ظاهر و حرکت اشیا هستند. مدل‌سازی حرکت، الگوهای حرکتی اشیا را در طول زمان بازنمایی می‌کند تا مسیرهای حرکتی آن‌ها به طور مؤثری ردیابی شده، موقعیت‌های آینده آن‌ها پیش‌بینی گردد و دقت الگوریتم‌های ردیابی اشیا افزایش یابد. از جمله روش‌های استاندارد برای پیش‌بینی خط سیر حرکتی می‌توان به فرض سرعت ثابت، فیلتر کالمن و مدل سازی احتمالاتی با استفاده از مدل‌های ترکیبی گوسی اشاره کرد. در سال‌های اخیر، تکنیک‌های یادگیری عمیق نیز معرفی شده‌اند که الگوهای حرکتی پیچیده را مستقیماً از داده‌ها مدل کرده و مسیرهای حرکتی معتبر تولید می‌کنند. از سوی دیگر مدل سازی ظاهر، امکان تمایز و ردیابی اشیا را بر اساس ویژگی‌های بصری آن‌ها فراهم می سازد. این نوع مدل‌سازی شامل استخراج و بازنمایی تغییرات ظاهری اشیا در طول فریم ها و مدیریت چالش هایی مانند تغییرات روشنایی، انسداد و تغییرات ظاهری شیء است (Psalta et al., 2024). شباهت ظاهری را می‌توان با استفاده از شباهت کسینوسی میان بردارهای ویژگی شناسایی مجدد اندازه‌گیری کرد (Y. Zhang et al., 2022).

## ۳- روش شناسی

توسعه تکنیک‌های بینایی کامپیوتری در دهه گذشته امکان استفاده قابل‌اطمینان از داده‌های بصری برای ارائه اطلاعات در زمینه حمل‌ونقل را فراهم کرده و باعث شده که استخراج خط سیر حرکتی و سایل نقلیه به مسئله‌ای قابل‌توجه تبدیل شود. در مقاله (Shepelev et al., 2020)، شپلف و همکاران داده های لحظه‌ای ترافیک را با استفاده از دوربین‌های نظارتی از زوایای دید مختلف جمع‌آوری می‌کنند، که منجر به بروز مسئله انسداد می شود. برای رفع این مشکل، بخش‌هایی از جاده که قبل یا بعد از شیء ثابت قرار دارند، ماسک می‌شود. همچنین، برای تشخیص وسایل نقلیه از YOLOv3 و به‌همراه آن از ردیاب SORT بهره گرفته شده است که مبتنی بر الگوریتم مجار ستانی و فیلتر کالمن است. در مقاله (Sharma et al., 2022)، شارما و همکاران یک شبکه عصبی پیچشی مبتنی بر ناحیه اصلاح شده را پیشنهاد می‌کنند. در روش پیشنهادی آن‌ها از فیلتر میانه مکانی اصلاح‌شده برای حذف نویز از فریم‌ها استفاده شده و شبکه از پیش آموزش‌دیده GoogleNet آموزش داده می‌شود. برای افزایش کارایی تشخیص، دوازده کادر پیش‌فرض به کار گرفته شده تا اشیای کوچک‌تر که در فاصله بیشتری از دوربین قرار دارند نیز در نظر گرفته شوند. شبکه پیشنهادی آن‌ها از الحاق ویژگی‌ها برای جلوگیری از ازدست‌رفتن اطلاعات در طول پردازش‌های عمیق به دلیل یکنواختی اندازه‌ها استفاده کرده و همچنین، روش نمونه‌گیری منفی نیز برای افزایش دقت شبکه به کار گرفته می‌شود. این روش برای غلبه بر شرایط نوری ضعیف در شب، به یک روش تبدیل صحنه تکیه دارد. در مطالعه (Song et al., 2019)، سانگ و همکارانش سامانه‌ای برای تشخیص و ردیابی و سایل نقلیه معرفی کرده‌اند که در آن از YOLOv3 برای تشخیص اشیا و ORB برای ردیابی آن‌ها بهره گرفته شده است. این الگوریتم به دقت رضایت‌بخشی دست یافته، اما از نظر زمان پردازش همچنان نیاز به بهبود دارد زیرا برای حذف نقاط نویزی کاذب، از پالایش نقاط ویژگی با الگوریتم RANSAC و برآورد ماتریس هموگرافی استفاده می شود. در مطالعه (Gomaa et al., 2019)، در ابتدا وسایل نقلیه با استفاده از یک طبقه‌بند مبتنی بر شبکه عصبی پیچشی همراه با برچسب‌گذاری اجزای متصل شناسایی می شوند. سپس، ویژگی‌های حرکتی و سایل نقلیه تحلیل شده تا نویز حذف و و سایل نقلیه خوشه‌بندی شوند. به‌منظور بهبود نتایج، از راهبرد ترکیبی الگوریتم خوشه‌بندی K-means و اطلاعات ردیابی Optical Flow بهره گرفته شده است. در نهایت، روشی برای تخصیص وسایل نقلیه شناسایی‌شده به خوشه متناظر آن‌ها پیشنهاد شده تا با درنظرگرفتن ناحیه تلاقی میان نقاط شناسایی و ردیابی‌شده، از شمارش تکراری جلوگیری شود. اگرچه این الگوریتم به نتایج امیدوارکننده‌ای دست یافته است، اما زمان پردازش همچنان زیاد است.

یافته‌ها نشـــان می‌دهد که اکثر مطالعات پیشـــین قادر به پردازش بلادرنگ نیســـتند که موجب عدم توانایی در پاسخگویی به نیازهای فوری و پویا در سامانه‌های نظارتی می‌شـــود. علاوه‌براین، بســـیاری از این مطالعات فاقد ارزیابی‌های جامع و دقیق با استفاده از مجموعه‌ای متنوع از معیارهای ارزیابی هســـتند. روش‌های اســـتفاده‌شـــده در این پژوهش‌ها اغلب از دقت لازم برخوردار نبوده و توانایی انعکاس شرایط پیچیده و متغیر دنیای واقعی را ندارند. بدون ارزیابی دقیق و اصـــولی بر اســـاس داده‌های واقعی و صـــحنه‌هـای چـالش‌برانگیز نمی‌توان بـه طور قطعی در خصوص کارایی و بهینگی روش‌های پیشنهادی در شرایط عملی اظهارنظر کرد. در مقابل، چارچوب پیشنهادی در این تحقیق با هدف بهبود نقاط ضـعف ذکرشـده، به اسـتخراج خط سیر حرکتی صـحیح و پیوسـته وسـایل نقلیه از طریق تشـخیص و ردیابی بلادرنگ اشـیا می‌پردازد. همچنین، بر اسـاس معیارهای ارزیابی مناسـب و کاربردی مورد ارزیابی قرار می‌گیرد تا نقاط قوت و ضـعف آن به طور شـفاف و مستند شناسایی و تحلیل گردد.

## ٤-روش اجرای تحقیق

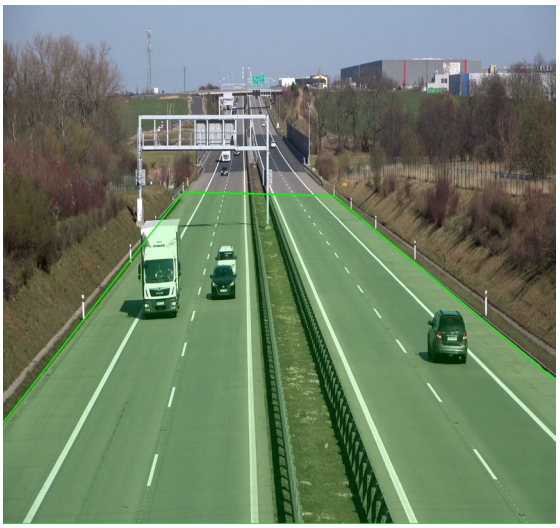
در این بخش، مؤلفه‌های کلیدی چارچوب پیشـــنهادی و جزئیات پیاده‌ســـازی آن‌ها به طور کامل و دقیق شـــرح داده می‌شود. چارچوب پیشنهادی شامل سه بخش اصلی است: تشخیص وسایل نقلیه، ردیابی وسیله نقلیه و استخراج خط ســـیر حرکتی. برای پیاده‌ســـازی این چارچوب، از زبان برنامه‌نویسی پایتون (نسخه ۳,۱۰,۱۲) و برای آموزش مدل و اجرای آزمایش‌هـای مختلف در این تحقیق، از واحـد پردازنده گرافیکی NVIDIA Tesla T4 اسـتفاده شـده اســـت. در ســـال‌های اخیر، کاربردهای یادگیری عمیق در داخل و خارج از حوزه بینایی کامپیوتری افزایش یافته اســـت. داده‌ها به‌عنوان عامل محدودکننده و محرک اصـــلی تأثیرگذار بر آموزش شبکه‌های عصبی عمیق شناخته شده‌اند (Ogunrinde, 2023). حتی بهترین معمـاری‌هـا ممکن است در صورت آموزش بر روی مجموعه‌داده‌ای که الزامات را برآورده نمی‌کند، به مدلی ضـعیف تبدیل شـوند. یکی از الزامات کلیدی این است که مجموعه‌داده باید نماینده باشد، به این معنی که تا حد امکان، حالات ممکن را پوشش دهد (Sousa, 2023). فرایند انتخاب و برچسب‌گذاری فریم‌ها نقش اساسی در عملکرد مدل دارد، زیرا دقت و کارایی مدل به کیفیت و تناســـب داده‌های آموزشـــی آن وابســـته اســـت (Christián, 2023). با هدف تأمین داده‌های آموزشـــی مناســـب، مجموعه‌داده‌ای تحت عنوان Surveillance Cameras (Bayat, 2026) تهیه شده است.

### ٤-۱-مجموعه‌داده Surveillance Cameras

مجموعه‌داده Surveillance Cameras شـــامل ۷۵٤۰۰ برچسب از ۵۰۳۵ تصویر ترافیکی است که در دو نسخه بدون افزایش داده و با افزایش داده تهیه شـــده اســـت. این مجموعه‌داده از ۳۵۲۵ تصویر آموزشی (۷۰ درصد)، ۱۰۰۷ تصویر اعتبار سنجی (۲۰ درصد) و ۵۰۳ تصویر آزمایشی (۱۰ درصد) تشکیل شده است و نحوه‌ی جداسازی داده‌ها به صورت کاملاً تصادفی بوده است. همچنین، شامل سه دســـته مختلف از وســـایل نقلیه بوده و توزیع یکنواخت نمونه‌های وسـایل نقلیه در هر سـه زیرمجموعه‌داده رعایت شـده اسـت. حدود ۹٤ درصـد از وسـایل نقلیه را خودرو تشکیل می‌دهد و ٦ درصد باقی‌مانده بین اتوبوس و کامیون تقسیم شده است. تصاویر این مجموعه‌داده از پلتفرم‌های مختلف مـانـنـد Kaggle، RoboFlow و Youtube با رزولوشن‌های متفاوت جمع‌آوری شده و شامل ۲۵ صحنه ترافیکی مختلف، نظیر تقاطع‌ها، جاده‌های محلی، راه‌های شـــریانی و بزرگراه‌ها، از دیدگاه‌ها و ارتفاعات مختلف می‌شـــود. هـدف اصـــلی از تهیـه این مجموعـه‌داده، برچسب‌گذاری باکیفیت و اصلاح برچسب‌های بی‌کیفیت است که بخش قابل توجهی از پیکسل‌های تصویر را شامل می‌شوند. ضروری است که کادر محاطی به نزدیک‌ترین حد ممکن به شیء کشیده شود و مرزهای آن را مشخص کند. به طور میانگین در هر تصـــویر ۱۵ برچســـب وجود دارد. در نســـخـه بـا افزایش داده، رویکردهـای متنوعی نظیر معکوس‌ســـازی افقی، چرخش، روشـــنایی، نوردهی و نویز تصادفی، برای افزایش داده در سطح تصویر در مجموعه‌داده آموزشی اعمال شده است.

### ٤-٢-تشخیص وسایل نقلیه

تشخیص وسایل نقلیه در صحنه‌های ترافیکی اهمیت بسیاری در پایش و استخراج خط سیر حرکتی وسایل نقلیه دارد (Mehrizi et al., 2021). در رویکرد ردیابی مبتنی بر تشخیص، عملکرد مدل تشخیص اشیا به طور مستقیم بر روی فرایند و نتایج کلی ردیابی تاثیر می‌گذارد (Messoussi et al., 2021). برای اهداف این پژوهش چندین الزام وجود دارد. علاوه بر نیاز عمومی برای تمام شناساگرها یعنی بالاترین دقت ممکن، مدل باید قادر به اجرای بلادرنگ نیز باشد. بنابراین، انتخاب الگوریتمی که میان بار محاسباتی و دقت تشخیص تعادل برقرار کند، ضروری است. در مرحله تشخیص وسایل نقلیه، از YOLOv10 که یک شناساگر اشیا پیشرفته مبتنی بر یادگیری عمیق است، استفاده می‌شود. YOLO مخفف عبارت (فقط یک بار نگاه کن) است که به نحوه عملکرد این الگوریتم اشاره داشته (Emiyah, 2021) و به دلیل سرعت بالا و عملکرد رقابتی شناخته می‌شود، اما برای دستیابی به زمان تشخیص سریع‌تر، مقداری از دقت را فدا می‌کند، که آن را به گزینه‌ای مناسب برای کاربردهای بلادرنگ که سرعت در آن‌ها اولویت دارد، تبدیل می‌کند (O. Boadu, 2024). آزمایش‌های گسترده روی بنچمارک‌های استاندارد تشخیص اشیا، نشان می‌دهند که YOLOv10 می‌تواند به طور قابل‌توجهی عملکرد بهتری نسبت به مدل‌های پیشرفته پیشین در زمینه تعادل میان دقت و محاسبات در مقیاس های مختلف مدل ارائه دهد (Wang et al., 2024). علاوه بر دقت و سرعت پردازش، انعطاف‌پذیری و سرعت زیاد توسعه آن از جمله دلایلی است که این مدل به‌عنوان مدل تشخیص وسایل نقلیه انتخاب شده است. YOLOv10 در شش مقیاس مختلف ارائه شده که از نظر عمق و عرض با یکدیگر تفاوت دارند، درحالی‌که ساختار کلی مدل حفظ می‌شود و با افزایش اندازه، پیچیدگی آن نیز افزایش می‌یابد. در راستای دستیابی به تعادلی بهینه میان بار محاسباتی و دقت تشخیص، مدل YOLOv10m (سایز متوسط) برای چارچوب پیشنهادی انتخاب شده است. خروجی این مدل شامل مختصات کادر محاطی، دسته و امتیاز اطمینان است که در قالب آرایه‌های عددی، مکان دقیق هر وسیله نقلیه و نوع آن را در هر فریم مشخص می سازد. سایر جزئیات مربوط به YOLOv10 در پیوست الف ارائه شده است. برای بهبود عملکرد و کاهش زمان آموزش از طریق تسریع همگرایی مدل، از وزن‌هایی که از پیش بر روی مجموعه‌داده MS COCO آموزش داده شده‌اند استفاده می‌شود. با بهره‌گیری از وزن‌های از پیش آموزش‌دیده به‌عنوان وزن‌های اولیه مدل، YOLOv10m بر روی مجموعه‌داده Surveillance Cameras تنظیم دقیق شده و مدل در طی ۲۰۰ دوره و با اندازه دسته ١٦ آموزش داده شده است. فرایند آموزش شامل به‌روزرسانی وزن‌های مدل به‌صورت تدریجی بر اساس میزان خطای آموزش و اعتبارسنجی است. در پروسه پیشنهادی، تشخیص‌هایی که امتیاز اطمینان آن‌ها کمتر از آستانه از پیش تعیین‌شده (معادل ٠,٤) بود، فیلتر شدند تا عملکرد شناسایی بهتری حاصل شود. همچنین، باتوجه به ارتفاع زیاد دوربین های ترافیکی و زاویه دید گسترده، برای افزایش دقت تشخیص و سرعت پردازش، یک سطح مقطع مناسب که دوربین دید مطلوبی داشته باشد، انتخاب می‌شود. ناحیه شناسایی به‌صورت یک ناحیه چند ضلعی در دستگاه مختصات تصویر تعریف می شود (شکل ۳).



**شکل ۳. تنظیم ناحیه شناسایی**

### ٤-٣-ردیابی وسایل نقلیه

گام بعدی شامل به‌کارگیری یک روش ردیابی اشیا به‌منظور استخراج خط سیر حرکتی وسایل نقلیه است. در این راستا، از رویکرد ردیابی مبتنی بر تشخیص استفاده شده که در آن

از الگوریتم ByteTrack برای ردیابی وسایل نقلیه در محدوده تشخیص بهره گرفته می‌شود. با استفاده از الگوریتم ردیابی ByteTrack که به طور مؤثر اشیا را با درنظرگرفتن تمامی کادرهای تشخیص (کادرهای تشخیص با امتیاز بالا و پایین) ردیابی می‌کند، تطبیق میان فریم‌های متوالی برقرار می‌شود. این الگوریتم با تخصیص شناسه‌های منحصربه‌فرد به هر وسیله نقلیه و ذخیره موقعیت آن در فریم‌های مختلف، امکان ردیابی پیوسته و دقیق را فراهم می‌آورد. ByteTrack تغییرات شناسه بسیار کمی دارد که نشان‌دهنده اثربخشی بالای تطبیق همه کادرهای تشخیص در شرایط انسداد است. این الگوریتم پایداری ردیابی در فریم‌ها را حتی در مواجهه با چالش‌هایی مانند انسداد و تداخلات پیچیده تضمین کرده و سرعت اجرای قابل‌توجهی را بدون کاهش دقت ارائه می‌دهد که آن را به گزینه‌ای مناسب برای کاربردهای بلادرنگ تبدیل می‌کند (Katariya, 2023; Y. Zhang et al., 2022).

سایر جزئیات مربوط به الگوریتم ردیابی ByteTrack در پیوست ب ارائه شده است. پارامترهای آستانه فعال‌سازی مسیر ردیابی، حداقل آستانه تطابق، حداقل تعداد فریم‌های متوالی و حافظه موقت مسیرهای ازدست‌رفته بر عملکرد ردیاب ByteTrack تأثیر گذاشته و باید به‌صورت مشترک تنظیم شوند. آستانه اطمینان تشخیص برای فعال‌سازی مسیر ردیابی توسط پارامتر آستانه فعال‌سازی مسیر ردیابی و آستانه تطابق مسیرهای ردیابی با تشخیص‌ها توسط پارامتر حداقل آستانه تطابق تعیین می‌شود. پارامتر حداقل تعداد فریم‌های متوالی برای تأیید یا رد اعتبار فرایند ردیابی استفاده می‌شود و بیانگر تعداد فریم‌های متوالی است که یک شیء باید ردیابی شود تا به‌عنوان یک مسیر ردیابی معتبر در نظر گرفته شود و باید مقداری برای آن انتخاب شود که بیشترین تعداد مسیرهای حرکتی نادرست را از تحلیل رد کند. اگر فرایند ردیابی با یک تشخیص نادرست شروع شود، احتمال کمی وجود دارد که تعداد زیادی کادر دیگر با آن موقعیت نادرست اولیه مطابقت پیدا کنند. در اینجا است که پارامتر حداقل تعداد فریم‌های متوالی می‌تواند برای تشخیص میان ردیابی‌ای که با یک موقعیت نادرست شروع شده و فرایند ردیابی معتبر استفاده شود. آستانه برای حفظ مسیرهای ردیابی ازدست‌رفته توسط حافظه موقت مسیرهای ازدست‌رفته کنترل می‌شود. در پروسه پیشنهادی، مقدار پیش‌فرض آستانه فعال‌سازی مسیر ردیابی برابر با ۰,۵ در نظر گرفته شده است. در مرحله تخصیص خطی، اگر نسبت اشتراک به اجتماع بین کادر تشخیص و کادر مسیر ردیابی کمتر از ۰/۷ باشد، تطبیق رد خواهد شد. هر شیء باید حداقل در ۲ فریم متوالی ردیابی شود تا به‌عنوان یک مسیر ردیابی معتبر در نظر گرفته شود. مسیرهای ردیابی ازدست‌رفته به مدت ۲۵ فریم نگهداری می‌شوند تا در صورت ظاهرشدن مجدد، بازیابی شوند. تا زمانی که یکی از دو شرط زیر نقض شود، ردیابی ادامه می‌یابد: شرط اول زمانی رخ می‌دهد که مسیر ردیابی به مدت تعداد فریم‌های برابر با پارامتر حافظه موقت مسیرهای ازدست‌رفته بدون تطبیق باقی بماند. شرط دوم زمانی است که نقطهٔ میانی در لبهٔ پایین کادر تشخیص از آخرین موقعیت خط سیر حرکتی از مرزهای ناحیه شناسایی خارج شود. در این نقطه، فرض می‌شود که شیء از صحنه خارج شده و فرایند ردیابی خاتمه می‌یابد.

### ٤-٤-استخراج خط سیر حرکتی وسایل نقلیه

پس از پردازش داده‌های ویدئویی از طریق اجرای مدل‌های تشخیص و ردیابی اشیا، اطلاعات وسایل نقلیه از جمله دسته، شناسه، جهت حرکت و مختصات کادر محاطی استخراج شده و دنباله‌ای از نقاط یا کادرهای محاطی به دست می‌آید که توالی مکان یک وسیلهٔ نقلیه در فریم‌های مختلف را نشان می‌دهد و این دنباله به‌عنوان خط سیر حرکتی وسایل نقلیه تعریف می‌شود. به‌طورکلی، عملکرد این چارچوب متکی بر تشخیص اشیا است که به طور مستقیم بر نتایج ردیابی و کیفیت استخراج خط سیر حرکتی وسایل نقلیه تأثیر می‌گذارد. استخراج دقیق خط سیر حرکتی وسایل نقلیه، امکان دستیابی به اطلاعات دقیق سطح بالایی از جریان ترافیک را فراهم می‌سازد، اما این امر کاملاً وابسته به داده‌های سطح پایین ترافیک مانند اطلاعات تشخیص و ردیابی وسایل نقلیه است (Azimjonov & Özmen, 2022).

**٤-٥-ملاحظات اخلاقی**

اخلاق هوش مصنوعی به ابهامات و پیامدهای اخلاقی مربوط به توسعه و پیاده‌سازی فناوری‌های هوش مصنوعی اشاره دارد. این امر شامل درنظرگرفتن نگرانی‌ها و معضلات اخلاقی مرتبط با هوش مصنوعی است. بنابراین، ضروری است که بر هرگونه ملاحظات اخلاقی که این مطالعه باید با آن‌ها مطابقت داشته باشد، تأکید شود. جدول ۱ برخی از این ملاحظات را نشان می‌دهد (Christián, 2023). بر اساس دستورالعمل‌های عمومی در زمینه اخلاق هوش مصنوعی، این مطالعه مسائل مرتبط با دقت، شفافیت، انصاف و سوگیری و حریم خصوصی و امنیت را مدنظر قرار داده است.

دقت: مجموعه‌داده تهیه‌شده در این پژوهش بر شرایط عادی تمرکز دارد، بنابراین ممکن است نتوان آن را به سایر مناطق با شرایط آب‌وهوایی خاص تعمیم داد. تمرکز اصلی بر کیفیت و دقت برچسب‌گذاری داده‌ها است که برای آموزش الگوریتم‌های تشخیص اشیا ضروری است. برچسب‌گذاری نادرست یا نقص در حاشیه‌نویسی می‌تواند منجر به عملکرد غیرکارآمد مدل شود.

شفافیت: توصیف واضح و دقیقی از نحوه جمع‌آوری و تهیه داده‌ها در مجموعه‌داده و چگونگی آموزش مدل وجود دارد تا امکان بررسی و اعتبارسنجی قابل‌اتکا توسط جامعه علمی فراهم شود. این مطالعه از الگوریتم‌های یادگیری عمیق برای تشخیص و ردیابی وسایل نقلیه استفاده کرده و نحوه عملکرد آن‌ها به‌خوبی مستند شده تا اطمینان حاصل شود که به‌درستی عمل می‌کنند و هیچ‌گونه رفتار ناخواسته‌ای از خود نشان نمی‌دهند. علاوه‌براین، مجموعه‌داده به طور عمومی در دسترس قرار گرفته است.

انصاف و سوگیری: سوگیری در داده‌های آموزشی ممکن است منجر به نتایج تحریف‌شده یا تبعیض‌آمیز شود که اهمیت استفاده از داده‌های متنوع و نماینده را برجسته می‌کند. درنتیجه، داده‌ها از طیف وسیعی از صحنه‌های ترافیکی مختلف جمع‌آوری شده است. علاوه‌براین، از مجموعه‌داده‌های متفاوتی برای آموزش مدل و ارزیابی کیفی چارچوب پیشنهادی استفاده شده است.

حریم خصوصی و امنیت: در مجموعه‌داده تهیه شده، چهره رانندگان و سایر افراد قابل شناسایی بصری نیست و هیچ رویداد غیرمعمول (مانند تصادف) رخ نمی‌دهد. همچنین صحنه‌ها فاقد هرگونه نشانه مکانی متمایز (از جمله نماهای شاخص ساختمان‌ها/لوگوها و یا اطلاعات جانبی حاوی مختصات جغرافیایی) هستند.

## تجزیه‌وتحلیل یافته‌های تحقیق

بخش حاضر به تحلیل جامع و دقیق نتایج حاصل از پیاده‌سازی و ارزیابی چارچوب پیشنهادی در زمینه استخراج خط سیر حرکتی وسایل نقلیه می‌پردازد. در راستای ارزیابی عملکرد مدل تشخیص اشیا، از مجموعه‌داده آزمایشی Surveillance Cameras استفاده شده است. این مجموعه‌داده شامل صحنه‌های شلوغ با موارد متعددی از انسدادهای جزئی و کامل موقتی است که سناریوهای پیچیده‌تری برای تشخیص اشیا ایجاد کرده و می‌تواند برای نشان‌دادن اثربخشی مدل پیشنهادی در شرایط محیطی چالش‌برانگیز از نظر تراکم و انسداد مفید واقع شود.

### معیارهای ارزیابی

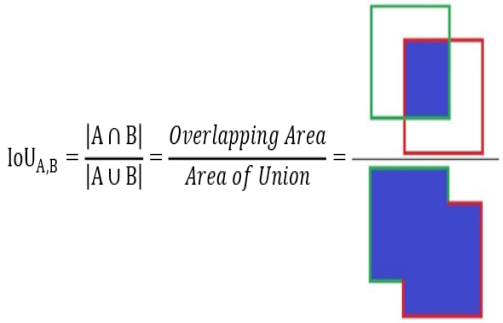
ارزیابی دقیق و اصولی بر اساس معیارهای مشخص، نه‌تنها به شفافیت کمک می‌کند، بلکه امکان شناسایی نقاط قوت و ضعف را نیز فراهم می‌سازد. معیارهای ارزیابی زیر برای سنجش کارایی مدل‌ها در وظیفه تشخیص اشیا به کار می‌روند:

**جدول ۱. اصول کلیدی قابلیت اطمینان و اخلاق در سامانه‌های هوش مصنوعی**

| اصل | تعریف |
|---|---|
| انصاف | به اصلی اشاره دارد که سامانه‌های هوش مصنوعی باید در فرایندهای تصمیم‌گیری خود بی‌طرف و بدون سوگیری عمل کنند. |
| شفافیت | به توانایی درک و توضیح اینکه چرا و چگونه یک سامانه تصمیم خاصی اتخاذ کرده یا به روشی خاص عمل کرده است، اشاره دارد. |
| پاسخ‌گویی | به مفهوم مسئول دانستن افراد یا نهادها برای اقدامات و تصمیماتی که توسط هوش مصنوعی اتخاذ شده، اشاره دارد. |
| کنترل‌پذیری | به توانایی مدیریت و نظارت بر رفتار و اقدامات هوش مصنوعی اشاره دارد. |

ماتریس تداخل برای بررسی عملکرد دسته‌بندی مدل و شناسایی اشتباهات خاصی که بر هر دسته تأثیر می‌گذارند، استفاده شده و شامل چهار مؤلفه است که در جدول ۲ شرح داده شده‌اند (S. Zhang, 2021). این مؤلفه‌ها برای ارزیابی دقت و قابلیت اطمینان پیش‌بینی‌های مدل به کار می‌روند.

نسبت اشتراک به اجتماع (شکل ٤)، میزان تطابق میان کادر محاطی پیش‌بینی‌شده و کادر محاطی واقعی را اندازه‌گیری می‌کند (Christián, 2023; Sousa, 2023; Stryk, 2023) و مقدار آن بین ۰ و ۱ متغیر است که در آن مقدار ۱ به معنی تطابق کامل و مقدار ۰ به معنی عدم اشتراک بین آن‌ها است (Sousa, 2023).




**شکل ٤. محاسبه نسبت اشتراک به اجتماع**

این نسبت برای تعیین صحت تشخیص بر اساس یک آستانه از پیش تعیین‌شده به کار می‌رود. اگر مقدار این نسبت برابر یا بیشتر از آستانه تعیین شده باشد، آن را مثبت واقعی می‌نامیم. اما اگر این آستانه را برآورده نکند، آن را مثبت کاذب در نظر می‌گیریم (Christián, 2023; Stryk, 2023). همچنین، اگر هیچ پیش‌بینی‌ای با نسبت برابر یا بیشتر از آستانه برای یک برچسب واقعی وجود نداشته باشد، به‌عنوان منفی کاذب در نظر گرفته می‌شود که به معنای تشخیص ازدست‌رفته است. با تغییر آستانه، پژوهشگران می‌توانند سطح سخت‌گیری ارزیابی را کنترل و مطابق با نیاز خود تنظیم کنند: آستانه بالاتر دقت عددی کمتری ارائه داده اما دقیق‌تر است، درحالی‌که آستانه پایین‌تر دقت عددی بیشتری ارائه داده و در پذیرش پیش‌بینی‌ها انعطاف‌پذیرتر است (Christián, 2023).

دقت، نسبت پیش‌بینی‌های مثبت واقعی را از مجموع تمام پیش‌بینی‌های مثبت انجام‌شده (که شامل مثبت‌های واقعی و مثبت‌های کاذب است) می‌سنجد:

$$\text{Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}} \quad (۳)$$

از سوی دیگر، بازیابی نسبت تعداد وسایل نقلیه‌ای است که به‌درستی شناسایی شده‌اند (مثبت‌های واقعی)، به کل وسایل نقلیه‌ای که باید شناسایی می‌شدند (شامل مثبت‌های واقعی و منفی‌های کاذب) (Gomaa et al., 2022; Zhou et al., 2021):

$$\text{Recall} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}} \quad (٤)$$

پس از محاسبه مقادیر دقت و بازیابی، می‌توان منحنی دقت – بازیابی را ترسیم کرده و از مساحت زیر منحنی به‌عنوان معیار عملکرد استفاده کرد. منحنی دقت – بازیابی با تغییر آستانه اطمینان کادرهای محاطی پیش‌بینی‌شده ترسیم می‌شود (Psalta et al., 2024; Stryk, 2023). دقت متوسط با میانگین‌گیری از مقادیر دقت در تمام مقادیر بازیابی از ۰ تا ۱ محاسبه می‌شود (Christián, 2023). این معیار مساحت زیر منحنی دقت – بازیابی را محاسبه کرده و یک ارزیابی کلی از دقت تشخیص ارائه می‌دهد (Christián, 2023; Sousa, 2023; Stryk, 2023):

**جدول ۲. توصیف مؤلفه‌های ماتریس تداخل**

| مؤلفه | توضیحات |
|---|---|
| مثبت واقعی | برچسب واقعی که به طور صحیح شناسایی شده است. |
| مثبت کاذب | کادر محاطی که به‌اشتباه شناسایی شده و شیء وجود ندارد یا تشخیص نادرستی از یک شیء موجود طبق برچسب واقعی است. |
| منفی کاذب | برچسب واقعی که شناسایی نشده است. |
| منفی واقعی | در وظیفه تشخیص اشیا، ممکن است تعداد نامحدودی اشیا وجود داشته باشند که نباید در هیچ تصویری شناسایی شوند. |

$$AP = \int_{r=0}^{1} p(r)dr = \sum_{k=0}^{k} (r(k) - r(k+1)) \times p(k) \quad (۵)$$

که در رابطه فوق $p(r)$ منحنی دقت - بازیابی، $K$ تعداد آستانه‌ها، $(r(k) - r(k+1))$ تفاوت مقادیر بازیابی متناظر با آستانه‌های $k$ و $k+1$ و $p(k)$ مقدار دقت متناظر با آستانه $k$ است. میانگین دقت متوسط یک معیار مبتنی بر دقت متوسط است. دقت متوسط برای هر دسته به طور مستقل محاسبه می شود. سپس میانگینِ دقت‌های متوسط محاسبه‌شده برای تمامی دسته‌ها به‌عنوان میانگین دقت متوسط گزارش می‌شود (Patil, 2021; Sousa, 2023):

$$\text{mean Average Precision} = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (۶)$$

که در آن $AP_i$ دقت متوسط برای دسته $i$ام و $N$ تعداد کل دسته‌ها است. در محاسبه میانگین دقت متوسط، ارزیابی معمولاً بر اساس یک یا چند آستانه نسبت اشتراک به اجتماع انجام می شود. این آستانه می‌تواند یک مقدار واحد یا مجموعه‌ای از مقادیر در یک بازه باشد. زمانی که یک مقدار واحد در نظر گرفته می شود، شناساگر تنها بر اساس همان آستانه ارزیابی می‌شود. در مقابل، زمانی که مجموعه‌ای از آستانه‌ها در یک بازه استفاده می‌شود، عملکرد مدل در چندین آستانه مختلف ارزیابی شده و نتیجه به‌صورت جامع‌تری گزارش می‌شود (Sousa, 2023).

معیار ارزیابی امتیاز F1 بین ۰ و ۱ متغیر بوده و ارتباط نزدیکی با دقت و بازیابی دارد و میانگین هارمونیک این دو معیار را نشان می‌دهد (Heidari et al., 2022; Zhou et al., 2021):

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (۷)$$

مقایسه شناساگرها از نظر صرف زمان نیز اهمیت دارد. برای این منظور، از زمان استنتاج استفاده می‌شود که به مدت زمانی اطلاق می‌شود که مدل برای پردازش یک ورودی و تولید پیش‌بینی نیاز دارد که می‌تواند به صورت فریم بر ثانیه (تعداد تصاویری که می‌توان در هر ثانیه پردازش کرد) بیان شود (Stryk, 2023).

### ارزیابی عملکرد مدل تشخیص اشیا

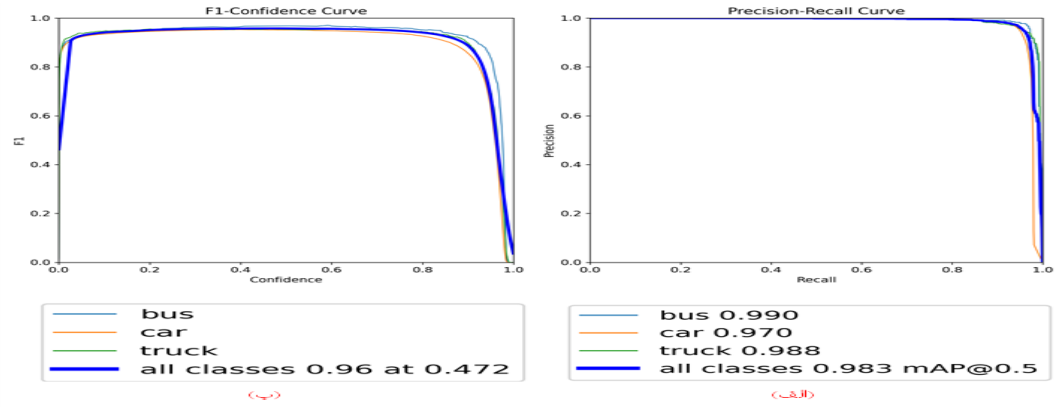
پس از اتمام مراحل آموزش و اعتبار سنجی، ارزیابی جامعی از عملکرد مدل تشخیص اشیا با استفاده از مجموعه‌ای متنوع از معیارهای ارزیابی صورت گرفت تا درک عمیق‌تری از قابلیت‌ها و محدودیت‌های آن فراهم سازد. هدف از این ارزیابی، ارائه معیاری دقیق از کارایی عملیاتی مدل برای کاربردهای احتمالی در دنیای واقعی است. در جدول ۳، عملکرد مدل تنظیم دقیق شده YOLOv10m بر روی مجموعه‌داده آزمایشی Surveillance Cameras گزارش شده است. معیارهای ارزیابی نظیر دقت، بازیابی، میانگین دقت متوسط و تأخیر محاسبه شده‌اند تا دقت و قابلیت اطمینان تشخیص مدل ارزیابی گردد. از نظر معیار دقت، مدل با دستیابی به ۹۶٫۳ درصد عملکرد مطلوبی از خود نشان داده است. این موضوع بیانگر توانایی مدل در شناسایی صحیح نمونه‌های مثبت در میان پیش‌بینی‌های مثبت خود است که در نهایت به افزایش دقت کلی در تشخیص منجر می‌شود. در معیار بازیابی، مدل به ۹۵٫۲ درصد دست یافت که بیانگر توانایی مدل در شناسایی درصد بالایی از نمونه‌های مثبت واقعی است که منجر به کاهش تعداد برچسب‌های واقعی شناسایی‌نشده و بازنمایی جامع‌تر اشیا مرتبط می‌شود که نشان می‌دهد بسیاری از اشیا مهم را شناسایی کرده و از دست نمی‌دهد. همچنین، مدل به مقدار میانگین دقت متوسط برابر با ۹۸٫۳ درصد دست یافت که بیانگر تعادل مطلوب میان دقت و بازیابی در سطوح مختلف آستانه اطمینان بوده و عملکرد مطلوب مدل را بیش‌ازپیش تأیید می‌کند. برای آستانه نسبت اشتراک به اجتماع برابر با ۰٫۵، نتایج بسیار مثبت است، اما برای درک کلی مدل، لازم است که نتایج برای آستانه‌های مختلف به دست آید. بنابراین، میانگین دقت متوسط برای آستانه‌های ۰٫۵ تا ۰٫۹۵ محاسبه شده و مقدار برابر با ۸۷٫۱ در صد حاصل گردید. بدیهی است که این مقدار به طور قابل توجهی نسبت به میانگین دقت متوسط با آستانه ۰٫۵ کاهش یابد، زیرا شامل آستانه‌های نسبت اشتراک به اجتماع سختگیرانه‌تری است که در آن‌ها سطح هم‌پوشانی میان کادرهای پیش‌بینی شده و

برچسب‌های واقعی باید بسیار زیاد و تقریباً کامل باشد تا پیش‌بینی‌ها به‌عنوان تشخیص صحیح در نظر گرفته شوند. یافته‌های دقیق‌تری با تمرکز بر تحلیل مقادیر میانگین دقت متوسط دسته‌های مختلف وسایل نقلیه شناسایی شده به دست می‌آید. نتایج، میانگین دقت متوسط بالاتری را برای اتوبوس‌ها نسبت به سایر دسته‌ها نشان می‌دهد که بیانگر توانایی برجسته مدل در تشخیص دقیق این دسته از وسایل نقلیه در میان سایر نمونه‌ها است. این موضوع، قابلیت اطمینان بالای مدل را در تفکیک صحیح میان اتوبوس‌ها و سایر دسته‌های وسیله نقلیه تأیید می‌کند. برای دسته‌های خودرو و کامیون، میانگین دقت متوسط نشان‌دهنده عملکرد رضایت‌بخش مدل در شناسایی این دسته‌هاست. بااین‌حال، گاهی اوقات مدل اشیای دیگر را به‌اشتباه به‌عنوان خودرو دسته‌بندی می‌کند و در مواجهه با اشکال و ویژگی‌های ظاهری مشابه خودروها با چالش‌هایی روبه‌رو است. دسته‌بندی نادرست خودروها احتمالاً به دلیل عدم تفکیک و درنظرگرفتن دسته مجزا برای وسایل نقلیه‌ای از جمله ون و وانت است. موضوعی که منجر به شناسایی اشتباه کامیون‌ها به‌عنوان خودرو می‌شود.

برای ارزیابی سیستماتیک مدل، تأخیر سراسری (زمان اجرای موتور TensorRT از تنسور ورودی آماده تا تولید خروجی نهایی مدل بدون احتساب بارگذاری تصویر، پیش‌پردازش و انتقال حافظه) با پیروی از پروتکل گزارش شده در YOLOv10 (Wang et al., 2024) اندازه‌گیری شده است. ارزیابی تأخیر مدل تنظیم دقیق شده YOLOv10m روی پردازنده گرافیکی NVIDIA Tesla T4 در محیط گوگل کولب با استفاده از TensorRT FP16 نشان می‌دهد که می‌تواند با میانگین تأخیر ۵,۰۸ میلی‌ثانیه به ازای هر تصویر اجرا شود که بیانگر عملکرد مطلوب مدل از منظر سرعت پردازش بوده و قابلیت آن را برای کاربردهای بلادرنگ برجسته می‌سازد. اختلاف میان مقدار به‌دست‌آمده و تأخیر گزارش‌شده در مقاله YOLOv10 می‌تواند ناشی از چند عامل باشد؛ الف) تفاوت نسخه‌های TensorRT، CUDA و درایور ب) تفاوت در تنظیمات ساخت موتور و روش‌های اجرایی انتخاب شده برای لایه‌ها توسط TensorRT ج) نوسانات شرایط اجرایی پردازنده گرافیکی در محیط گوگل کولب د) تفاوت پیکربندی مدل تنظیم‌دقیق‌شده، از جمله تعداد دسته‌ها یا ساختار خروجی. بنابراین اختلاف مشاهده شده احتمالاً ناشی از تفاوت محیط اجرا، پیکربندی مدل تنظیم دقیق شده و جزئیات بهینه‌سازی TensorRT است.

**جدول ۳.ارزیابی مدل تنظیم دقیق شده YOLOv10m بر روی مجموعه‌داده آزمایشی Surveillance Cameras به تفکیک هر دسته**

| مدل | تأخیر (میلی ثانیه) | دسته | دقت | بازیابی | میانگین دقت متوسط (۰,۵) | میانگین دقت متوسط (۰,۹۵-۰,۵) |
|---|---|---|---|---|---|---|
| YOLOv10m | ۵,۰۸ | مجموع | ۰,۹۶۳ | ۰,۹۵۲ | ۰,۹۸۳ | ۰,۸۷۱ |
| (سایز متوسط) | | اتوبوس | ۰,۹۶۵ | ۰,۹۶۹ | ۰,۹۹ | ۰,۹۱۴ |
| | | خودرو | ۰,۹۶۹ | ۰,۹۳۶ | ۰,۹۷ | ۰,۸۳۲ |
| | | کامیون | ۰,۹۵۶ | ۰,۹۵۲ | ۰,۹۸۸ | ۰,۸۶۵ |




**شکل ۵. ارزیابی مدل تنظیم دقیق شده YOLOv10m بر روی مجموعه‌داده آزمایشی Surveillance Cameras**
**(الف) منحنی دقت – بازیابی (ب) منحنی امتیاز F1 – اطمینان**

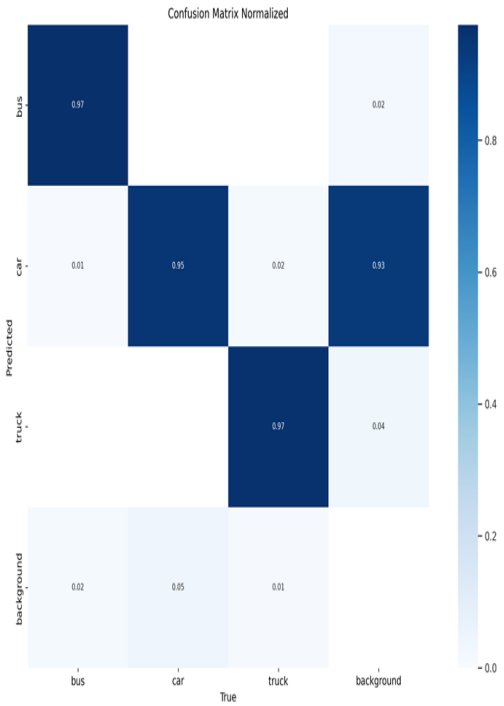
همان‌طور که در شکل ۵ (الف) نشان داده شده مساحت وسیع زیر منحنی دقت – بازیابی بیانگر میزان ثبات دقت در برابر افزایش تدریجی بازیابی است. امتیاز F1 نشان می‌دهد که مدل تنظیم دقیق شده YOLOv10m توانسته در سطوح پایین‌تر آستانه اطمینان، به سرعت به عملکرد مطلوب دست یابد و در سطوح بالاتر این عملکرد مطلوب را حفظ کرده و تعادلی بهینه میان دقت و بازیابی برقرار سازد و در نتیجه، در سناریوهای دشوار تشخیص نیز عملکرد دقیق‌تری ارائه دهد (شکل ۵ (ب)).

شکل ٦ توانایی مدل را در دسته‌بندی اشیا به تصویر می‌کشد. اکثر نمونه‌ها در هر دسته به‌درستی دسته‌بندی شده‌اند که نشان‌دهنده عملکرد خوب و پایدار مدل در تشخیص دسته‌های مختلف وسایل نقلیه است. بااین‌حال، همان‌طور که پیش‌تر اشاره شد، مشاهده می شود که گاهی اوقات کامیون و اتوبوس را به ترتیب با نرخ تشخیص نادرست معادل ۰,۰۲ و ۰,۰۱ به‌اشتباه به‌عنوان خودرو شناسایی می‌کند. با بررسی دقیق‌تر تصاویر مجموعه‌داده آزمایشی، مشخص شد که این خطای دسته‌بندی عمدتاً ناشی از شباهت ظاهری میان ون، وانت و کامیون است.

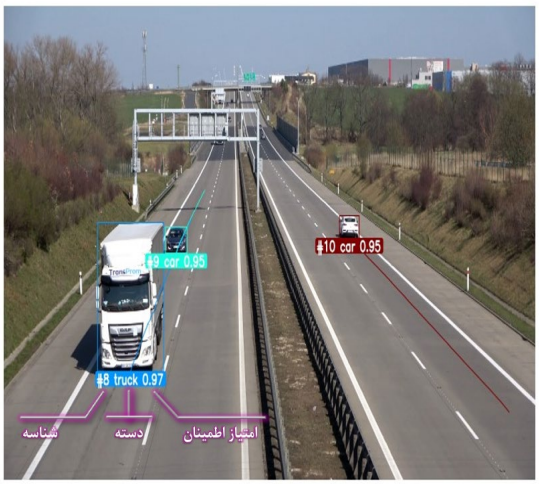
## خروجی بصری چارچوب پیشنهادی

خروجی چارچوب پیشنهادی در شکل ۷ نشان داده شده است. در این خروجی به طور پیش‌فرض، اطلاعاتی نظیر شناسه، دسته و امتیاز اطمینان وسایل نقلیه شناسایی شده در فریم‌های مختلف به همراه کادر محاطی و خط سیر حرکتی آن‌ها در فریم‌های متوالی نمایش داده شده و مسیر حرکتی شناسه‌ها به‌صورت خطوط پیوسته ترسیم می‌شود.

این خروجی اطلاعات دقیقی درباره تغییر مکان هر شناسه در طول زمان، ارائه داده و در تحلیل دقیق الگوهای حرکتی، پیش‌بینی مکان، استخراج پارامترهای ترافیکی و شناسایی تداخلات احتمالی کاربرد دارد.




**شکل ٦. ماتریس تداخل مدل تنظیم دقیق شده YOLOv10m بر روی مجموعه‌داده آزمایشی Surveillance Cameras**




**شکل ۷. خروجی بصری چارچوب پیشنهادی**

## ۵-نتیجه‌گیری

در این بخش پایانی، نتایج اصلی تحقیق مورد تجزیه‌وتحلیل قرار گرفته و بر اساس نتایج حاصل از پژوهش حاضر، مسیرهای تحقیقاتی آتی، چالش‌ها و محدودیت‌های موجود در این زمینه مورد بررسی قرار می‌گیرد. این پژوهش با هدف طراحی چارچوبی پیشرفته و کارآمد برای استخراج خط سیر حرکتی دقیق و پیوسته وسایل نقلیه از طریق تشخیص و ردیابی بلادرنگ اشیا در سامانه‌های نظارتی مبتنی بر تحلیل ویدئویی تدوین شده است. این رویکرد مبتنی بر بینایی کامپیوتری، فرایندی یکپارچه شامل تشخیص، ردیابی و استخراج خط سیر حرکتی وسایل نقلیه است. با هدف دستیابی به تعادلی بهینه میان معیارهای عملکردی و زمان پردازش در فرایند تشخیص و ردیابی وسایل نقلیه به‌منظور تضمین پاسخ‌دهی دقیق و به‌موقع، از دو الگوریتم YOLOv10 برای تشخیص اشیا و ByteTrack برای ردیابی آن‌ها بهره گرفته شده است.

این الگوریتم‌ها از نظر محاسباتی کارآمد بوده و امکان تشخیص و ردیابی بلادرنگ اشیا را با دقت بالا فراهم می‌سازند. با درک اهمیت دسترسی به مجموعه‌داده‌ای متنوع، باکیفیت و نماینده، مجموعه‌داده‌ای جدید تحت عنوان Surveillance Cameras با هدف برچسب گذاری باکیفیت از صحنه‌های مختلف ترافیکی، تهیه گردید. از تکنیک یادگیری انتقالی برای تنظیم دقیق مدل از پیش آموزش‌دیده YOLOv10m بر روی این مجموعه‌داده بهره گرفته شده است. مدل تنظیم دقیق شده YOLOv10m توانسته در معیارهای دقت، بازیابی و میانگین دقت متوسط به ترتیب به مقادیر ۹۶,۳، ۹۵,۲ و ۹۸,۳ درصد دست یابد و با تأخیر حدود ۵ میلی‌ثانیه اجرا شود. نتایج حاصل نشان می‌دهد که چارچوب پیشنهادی قادر است اکثر وسایل نقلیه را به طور دقیق شناسایی و به‌درستی دسته‌بندی کرده و نرخ فریم بالایی را در کاربردهای بلادرنگ تضمین کند. این عملکرد، در رویکرد ردیابی مبتنی بر تشخیص، فرایند تطبیق داده را بهبود بخشیده و ضمن حفظ پیوستگی شناسه‌ها در فریم‌های متوالی، امکان استخراج خط سیر حرکتی دقیق، پایدار و پیوسته وسایل نقلیه را فراهم می‌سازد.

از دستاوردهای اصلی این تحقیق می‌توان به کارایی در تشخیص و ردیابی بلادرنگ اشیا و استخراج خط سیر حرکتی دقیق و پیوسته وسایل نقلیه اشاره کرد که به طور مؤثر تغییر مکان شناسه‌ها را در طول زمان به‌درستی ثبت می‌کند. این چارچوب توانایی این فناوری‌ها را در حل چالش‌های دنیای واقعی نشان می‌دهد و راه را برای توسعه و نوآوری بیشتر در این زمینه هموار می‌سازد. یکی از محدودیت‌ها در استفاده از مدل‌های مبتنی بر یادگیری عمیق، دستگاه سخت‌افزاری محاسباتی است. در حال حاضر مدل‌های مبتنی بر یادگیری عمیق اغلب بر روی پردازنده‌های گرافیکی پیاده سازی می شوند که هزینه بالایی دارد تا اطمینان حاصل شود که مدل می‌تواند با سرعت نسبتاً بالا آموزش ببیند و پیش‌بینی کند. برای استفاده از پیش‌بینی‌کننده در یک وضعیت واقعی، تضمین چنین مرکز محاسباتی بزرگی دشوار است، به‌ویژه برای یک شبکه جاده‌ای بزرگ. بااین‌حال، به دلیل نیاز روزافزون به پردازش فوری و درون دستگاهی و همچنین پیشرفت‌های سریع در فناوری سخت‌افزار، تمایل فزاینده‌ای به آموزش و اجرای

این مدل‌ها به طور مستقیم بر روی دستگاه‌های لبه به وجود آمده اســـت. همچنین، تحلیل تأثیرات محیطی بر فرایند تشخیص اشیا یکی از مسائل اساسی و چالش‌برانگیز است. عملکرد تشـــخیص چارچوب می‌تواند تحت تأثیر شـــرایط نوری متنوع، شرایط جاده و شرایط جوی شدید مانند برف ســـنگین و مه قرار گیرد. بنابراین، اســـتفاده از مدل‌های تشـــخیص اشـــیا به همراه روش‌هایی نظیر تطبیق دامنه برای استخراج خط سیر حرکتی و سایل نقلیه در شرایط نوری، محیطی و جوی چـالش‌برانگیز می‌توانـد در پژوهش‌هـای آینده مورد بررسی قرار گیرد.

## ٦-مراجع


-Abdel-Aty, M., Wang, Z., Zheng, O., & Abdelraouf, A. (2023). Advances and applications of computer vision techniques in vehicle trajectory generation and surrogate traffic safety indicators. *Accident Analysis & Prevention*, *191*, 107191.
**doi.org/10.1016/J.AAP.2023.107191**

-Ait Ouallane, A., Bakali, A., Bahnasse, A., Broumi, S., & Talea, M. (2022). Fusion of engineering insights and emerging trends: Intelligent urban traffic management system. *Information Fusion*, *88*, 218–248.
**doi.org/10.1016/J.INFFUS.2022.07.020**

-Azimjonov, J., & Özmen, A. (2021). A real-time vehicle detection and a novel vehicle tracking systems for estimating and monitoring traffic flow on highways. *Advanced Engineering Informatics*, *50*, 101393.
**doi.org/10.1016/J.AEI.2021.101393**

-Azimjonov, J., & Özmen, A. (2022). Vision-based vehicle tracking on highway traffic using bounding-box features to extract statistical information. *Computers & Electrical Engineering*, *97*, 107560.
**doi.org/10.1016/J.COMPELECENG.2021.107560**

-Bakirci, M. (2024). Enhancing vehicle detection in intelligent transportation systems via autonomous UAV platform and YOLOv8 integration. *Applied Soft Computing*, *164*, 112015.
**doi.org/10.1016/J.ASOC.2024.112015**

-Bayat, M. (2026, July 9). *Surveillance Cameras Dataset*.

-Bunaiya Prayudha, R. (2021). AddInsight-Vision An Intelligent Video System for Object Detection in Traffic Scenes and Vehicle Re-Identification.

-Christián, U. (2023). Real Time Vehicle Detection for Intelligent Transportation Systems.

-Derek Adam Gloudemans, B., Adams, D., Bayen, A., Biswas, G., & Laine, F. (2023). Interstate 24 Motion: A Multi-Camera Vehicle Tracking Instrument For Large-Scale Trajectory Extraction.

-Ding, B. M. (2023). Radar and Camera Fusion in Intelligent Transportation System.

-Emiyah, C. (2021). Framework For Extracting Relevant Transportation Metrics From Drone Surveillance Videos Using Computer Vision & Machine Learning.

-García-Aguilar, I., García-González, J., Medina, D., Luque-Baena, R. M., Domínguez, E., & López-Rubio, E. (2024). Detection of dangerously approaching vehicles over onboard cameras by speed estimation from apparent size. *Neurocomputing*, *567*, 127057.
**doi.org/10.1016/J.NEUCOM.2023.127057**

-Gomaa, A., Abdelwahab, M. M., Abo-Zahhad, M., Minematsu, T., & Taniguchi, R. I. (2019). Robust Vehicle Detection and Counting Algorithm Employing a Convolution Neural Network and Optical Flow. *Sensors 2019,* Vol. 19, 19(20), 4588—4589.
**doi.org/10.3390/S19204588**

-Gomaa, A., Minematsu, T., Abdelwahab, M. M., Abo-Zahhad, M., & Taniguchi, R. ichiro. (2022). Faster CNN-based vehicle detection and counting strategy for fixed camera scenes. *Multimedia Tools and Applications*, *81*(18), 25443–25471.
**doi.org/10.1007/S11042-022-12370-9/TABLES/9**

-Hassaballah, M., Kenk, M. A., Muhammad, K., & Minaee, S. (2021). Vehicle Detection and Tracking in Adverse Weather Using a Deep Learning Framework. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 4230–4242.
**doi.org/10.1109/TITS.2020.3014013**

-Heidari, A., Navimipour, N. J., & Unal, M. (2022). Applications of ML/DL in the management of smart cities and societies based on new trends in information technologies: A systematic literature review. *Sustainable Cities and Society*, *85*, 104089.
**doi.org/10.1016/J.SCS.2022.104089**

-Hislop-Lynch, & Richard, S. (2023). Capture, processing and analysis of vehicle trajectories from multiple unmanned-aerial-vehicles with computer-vision and artificial-intelligence [The University of Queensland]. **doi.org/10.14264/AB80CF8**

-Hnewa, M. (2023). Object Detection for Autonomous Systems Operating Under Challenging Conditions.

-Katariya, V. A. (2023). Advancing highway safety: embedded-edge ai for real-time applications.
**doi.org/10.1109/TITS.2022.3172015**

-Li, J., Xu, Z., Fu, L., Zhou, X., & Yu, H. (2021). Domain adaptation from daytime to nighttime: A situation-sensitive vehicle detection and traffic flow parameter estimation framework. *Transportation Research Part C: Emerging Technologies*, *124*, 102946. **doi.org/10.1016/J.TRC.2020.102946**

-Li, T., Bian, Z., Lei, H., Zuo, F., Yang, Y. T., Zhu, Q., Li, Z., & Ozbay, K. (2024). Multi-level traffic-responsive tilt camera surveillance through predictive

*European Transport Research Review*, *11*(1), 1–16. **doi.org/10.1186/S12544-019-0390-4/TABLES/5**
-Sousa, R. M. P. de. (2023). *Detection and Tracking of Red Traffic Light Infringements Using DNNs*.
-Stryk, F. (2023). *Vehicle Location Detection and Distribution from Camera Images*. **doi.org/10.1109/AVSS.2017.8078520**
-VAŇO, J. (2024). Traffic surveillance system. In *theses.czJ VAŇOtheses.cz*.
-Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). *YOLOv10: Real-Time End-to-End Object Detection*.
-Yang, H. (2023). *Customizing Trustworthy Machine Learning and Advanced Computing Methods for Cooperative and Equitable Traffic Systems*.
-Yang, H. (Frank), Cai, J., Liu, C., Ke, R., & Wang, Y. (2023). Cooperative multi-camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning. *Transportation Research Part C: Emerging Technologies*, *148*, 103982. **doi.org/10.1016/J.TRC.2022.103982**
-Zhang, S. (2021). Improving Pedestrian Safety Using Video Data, Surrogate Safety Measures and Deep Learning. In *Electronic Theses and Dissertations, 2020-2023*. https://stars.library.ucf.edu/etd2020/593
-Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object Tracking by Associating Every Detection Box. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13682 LNCS*, 1–21. **doi.org/10.1007/978-3-031-20047-2_1**
-Zhou, S., Ng, S. T., Yang, Y., & Xu, J. F. (2021). Integrating computer vision and traffic modeling for near-real-time signal timing optimization of multiple intersections. *Sustainable Cities and Society*, *68*, 102775. **doi.org/10.1016/J.SCS.2021.102775**
correlated online learning. *Transportation Research Part C: Emerging Technologies*, *167*, 104804. **doi.org/10.1016/J.TRC.2024.104804**
-Mattingly, A. (2023). *Deep Learning Ensembles for Lightweight Object Detection*.

-Mehrizi, A., Gholampoor, I., & Movahedian, H. (2021). Automatic Traffic Incident Detection Based on Monitoring Camera Image Processing System.
-Messoussi, O., Nicolescu, G., & Bilodeau, G. A. (2021). Real-Time Vehicle Detection and Tracking From Surveillance Cameras in Urban Scenes.
-Muhib, R. B., Ahmad, I. S., & Boufama, B. (2023). Deep Learning-Based Vehicle Classification. In *Lecture Notes in Networks and Systems: 814 LNNS*. **doi.org/10.1007/978-3-031-47451-4_18**
-Ngeni, F., Mwakalonge, J., & Siuhi, S. (2024). Solving traffic data occlusion problems in computer vision algorithms using DeepSORT and quantum computing. *Journal of Traffic and Transportation Engineering (English Edition)*, *11*(1), 1–15. **doi.org/10.1016/J.JTTE.2023.05.006**
-O. Boadu, B. (2024). Dual Vision: Enhancing Autonomous Navigation With AutoGuardian CNN and YOLO Object Detection Synergy.
-Ogunrinde, I. O. (2023). Multi-Sensor Fusion for Object Detection and Tracking under Foggy Weather Conditions.
-Patil, J. B. (2021). Camera Video Data Processing of Traffic and Development of Microsimulation Traffic SUMO Model for Roundabout.
-Psalta, A., Tsironis, V., & Karantzalos, K. (2024). Transformer-based assignment decision network for multiple object tracking. *Computer Vision and Image Understanding*, *241*, 103957. **doi.org/10.1016/J.CVIU.2024.103957**
-Rezaei, M., Azarmi, M., & Mir, F. M. P. (2023). 3D-Net: Monocular 3D object recognition for traffic monitoring. *Expert Systems with Applications*, *227*, 120253. **doi.org/10.1016/J.ESWA.2023.120253**
-Sharma, P., Singh, A., Singh, K. K., & Dhull, A. (2022). Vehicle identification using modified region based convolution network for intelligent transportation system. *Multimedia Tools and Applications*, *81*(24), 34893–34917. **doi.org/10.1007/S11042-020-10366-X/METRICS**
-Shepelev, V., Aliukov, S., Glushkov, A., & Shabiev, S. (2020). Identification of distinguishing characteristics of intersections based on statistical analysis and data from video cameras. *Journal of Big Data*, *7*(1), 1–23. **doi.org/10.1186/S40537-020-00324-7/TABLES/9**
-Siddique, A. (2023). *Downstream Task Self-Supervised Learning for Object Recognition and Tracking*.
-Song, H., Liang, H., Li, H., Dai, Z., & Yun, X. (2019). Vision-based vehicle detection and counting system using deep learning in highway scenes.

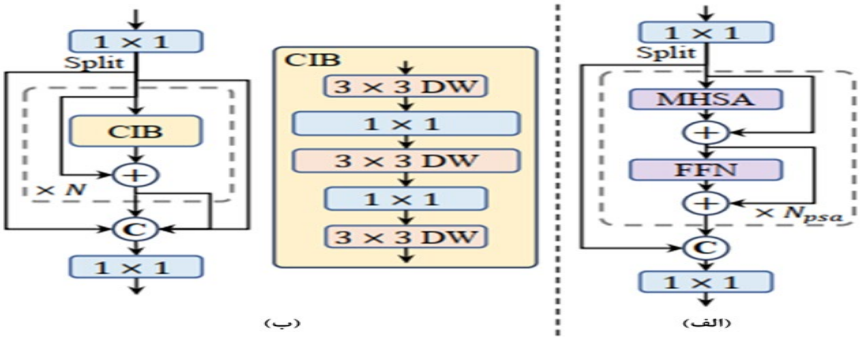
## پیوست الف

مدل‌های YOLOv10 با استفاده از بهینه‌ساز SGD و به مدت ۵۰۰ دوره آموزش داده شده‌اند. مقدار مومنتوم و نرخ کاهش وزن در SGD به ترتیب برابر با ۰٫۹۳۷ و $5 \times 10^{-4}$ تنظیم شده‌اند. نرخ یادگیری اولیه برابر با $1 \times 10^{-2}$ بوده که به‌صورت خطی تا مقدار $1 \times 10^{-4}$ کاهش می‌یابد. از تکنیک‌های افزایش داده Mosaic، Mixup و Copy-Paste استفاده شده و مدل‌ها بر روی ۸ کارت گرافیک NVIDIA 3090 آموزش داده شده‌اند. یک بخش پیش‌بینی اضافی مبتنی بر تطابق یک‌به‌یک به ساختار YOLOv10 افزوده شده است. این بخش دارای ساختاری مشابه با شاخه اصلی مبتنی بر تطابق یک‌به‌چند است و از اهداف بهینه‌سازی یکسانی بهره می‌برد، با این تفاوت که از تطابق یک‌به‌یک برای تعیین تخصیص برچسب‌ها بهره می‌گیرد. در طول آموزش، هر دو بخش پیش‌بینی به طور مشترک با مدل بهینه‌سازی شده و در زمان استنتاج، تنها از بخش پیش‌بینی یک‌به‌یک استفاده می‌شود. در تطابق یک‌به‌یک، از روش انتخاب بهترین تطابق استفاده می‌شود. در فرایند تخصیص، هر دو رویکرد از یک معیار برای ارزیابی کمی میزان انطباق بین پیش‌بینی‌ها و نمونه‌های واقعی استفاده می‌کنند. به‌منظور دستیابی به تطابق آگاه از پیش‌بینی در هر دو شاخه، از یک معیار تطابق واحد استفاده می‌شود:

$$m(\alpha, \beta) = s \cdot p^{\alpha} \cdot IoU(\hat{b}, b)^{\beta}$$

که در آن $p$ امتیاز دسته‌بندی، $\hat{b}$ و $b$ به ترتیب نشان‌دهنده کادر محاطی پیش‌بینی‌شده و نمونه واقعی و $s$ بیانگر پیش‌فرض مکانی است که مشخص می‌کند آیا نقطه مرجع پیش‌بینی درون نمونه قرار دارد یا خیر. فراپارامترهای $\alpha$ و $\beta$ تأثیر وظیفه پیش‌بینی معنایی و وظیفه رگرسیون موقعیت را متعادل می‌سازند. بخش دسته‌بندی YOLOv10 شامل دو کانولوشن تفکیک‌پذیر عمق‌محور با اندازه فیلتر ۳×۳ و یک کانولوشن ۱×۱ پس از آن است. همچنین، عملیات کاهش ابعاد مکانی و افزایش تعداد کانال‌ها از یکدیگر جدا شده و ابتدا از کانولوشن نقطه‌محور برای تنظیم بُعد کانالی و سپس از کانولوشن عمق‌محور برای نمونه‌کاهی مکانی بهره می‌گیرد. علاوه‌براین، یک طرح بلوک رتبه‌محور با بهره‌گیری از طراحی معماری فشرده ارائه می‌دهد. ساختار بلوک معکوس فشرده از کانولوشن‌های عمق‌محور کم‌هزینه برای ترکیب ویژگی‌ها در بُعد مکانی و کانولوشن‌های نقطه‌محور مقرون‌به‌صرفه برای ترکیب ویژگی‌ها در بُعد کانالی استفاده می‌کند (شکل ۸ (ب)). برای تخصیص بلوک‌های رتبه‌محور در یک مدل معین، تمامی مراحل آن بر اساس رتبه ذاتی هر مرحله به ترتیب صعودی مرتب می‌شود. سپس تغییر عملکرد مدل در صورت جایگزینی بلوک پایه در اولین مرحله پس از مرتب‌سازی با بلوک معکوس فشرده ارزیابی می‌گردد. اگر کاهش عملکردی نسبت به مدل اولیه مشاهده نشود، فرایند ادامه می‌یابد؛ در غیر این صورت، فرایند متوقف می‌شود. YOLOv10 در مدل‌هایی با مقیاس کوچک، از کانولوشن‌های عمق‌محور با فیلتر بزرگ درون ساختار بلوک معکوس فشرده در مراحل عمیق شبکه استفاده می‌کند. همچنین، از تکنیک پارامترسازی مجدد ساختاری استفاده می‌کند تا یک شاخه کانولوشن عمق‌محور ۳×۳ دیگر نیز اضافه شود.

علاوه‌براین، ماژول مؤثر خودتوجهی بخشی را ارائه می‌دهد که تنها در مرحله‌ای با کم‌ترین رزولوشن قرار داده می‌شود (شکل ۸ (الف)). ابعاد کوئری و کلید در ماژول خودتوجهی چندسر برابر با نصف ابعاد مقدار در نظر گرفته شده و به‌جای استفاده از نرمال‌سازی لایه‌ای از نرمال‌سازی دسته‌ای بهره گرفته می‌شود (Wang et al., 2024).




**شکل ۸ (الف) ماژول خودتوجهی بخشی (ب) بلوک معکوس فشرده**

**پیوست ب**

ورودی ByteTrack یک دنباله ویدئویی به همراه یک شناساگر اشیا است. یک آستانه برای امتیاز تشخیص $\tau$ تعیین می‌شود و خروجی آن مسیرهای ردیابی $\mathcal{T}$ است، که هر مسیر شامل کادر محاطی و شناسه شیء در هر فریم است. برای هر فریم، کادرهای تشخیص و امتیاز اطمینان آن‌ها با استفاده از شناساگر پیش‌بینی شده و بر اساس آستانه اطمینان تشخیص $\tau$ به دو دسته تقسیم می‌شوند: کادرهایی که امتیاز آن‌ها بالاتر از آستانه $\tau$ است، در مجموعه کادرهای تشخیص با امتیاز بالا $\mathcal{D}_{high}$ و کادرهایی که امتیاز آن‌ها کمتر از آستانه $\tau$ است، در مجموعه کادرهای تشخیص با امتیاز پایین $\mathcal{D}_{low}$ قرار داده می‌شوند. سپس، از فیلتر کالمن برای پیش‌بینی موقعیت جدید هر مسیر ردیابی در مجموعه $\mathcal{T}$ در فریم فعلی استفاده می شود. تطبیق اول با استفاده از معیار شباهت نسبت اشتراک به اجتماع میان کادرهای تشخیص با امتیاز بالا $\mathcal{D}_{high}$ و کادرهای پیش‌بینی‌شده مسیرهای ردیابی $\mathcal{T}$ (شامل تمام مسیرهای ردیابی از جمله مسیرهای ازدست‌رفته $\mathcal{T}_{lost}$) انجام می شود. سپس، برای انجام تطبیق بر اساس این شباهت، از الگوریتم مجارستانی استفاده شده و کادرهای تشخیص تطبیق نیافته در مجموعه $\mathcal{D}_{remain}$ و مسیرهای تطبیق نیافته در مجموعه $\mathcal{T}_{remain}$ نگهداری می‌شوند. تطبیق دوم میان کادرهای تشخیص با امتیاز پایین $\mathcal{D}_{low}$ و مسیرهای باقی‌مانده پس از تطبیق اول $\mathcal{T}_{remain}$ انجام می‌شود. مسیرهایی که مجدداً تطبیق نیافتن در مجموعه $\mathcal{T}_{re-remain}$ نگهداری شده و تمامی کادرهای تشخیص با امتیاز پایین که بدون تطبیق باقی مانده‌اند حذف می‌شوند. مسیرهای تطبیق نیافته پس از تطبیق دوم $\mathcal{T}_{re-remain}$، به مجموعه مسیرهای ازدست‌رفته $\mathcal{T}_{lost}$ منتقل می‌شوند. برای هر مسیر در $\mathcal{T}_{lost}$، تنها در صورتی که بیش از تعداد مشخصی فریم بدون تطبیق باقی بماند، از مجموعه مسیرهای ردیابی $\mathcal{T}$ حذف می‌شود. در نهایت، مسیرهای جدید از کادرهای تشخیص با امتیاز بالا که پس از تطبیق اول بدون تطبیق باقی مانده‌اند $\mathcal{D}_{remain}$، مقداردهی اولیه می‌شوند (Y. Zhang et al., 2022).

**الگوریتم ۱. شبه کد الگوریتم ByteTrack**

**Input:** A video sequence $V$; object detector Det; detection score threshold $\tau$
**Output:** Tracks $\mathcal{T}$ of the video

Initialization: $\mathcal{T} \leftarrow \emptyset$
**for** frame $f_k$ in $V$ **do**
  $\mathcal{D}_k \leftarrow \text{Det}(f_k)$ /* predict detection boxes & scores */
  $\mathcal{D}_{high} \leftarrow \emptyset$
  $\mathcal{D}_{low} \leftarrow \emptyset$
  **for** d in $\mathcal{D}_k$ **do**
    **if** d.score > $\tau$ **then**
      $\mathcal{D}_{high} \leftarrow \mathcal{D}_{high} \cup \{d\}$
    **End**
    **Else**
      $\mathcal{D}_{low} \leftarrow \mathcal{D}_{low} \cup \{d\}$
    **End**
  **End**
  **for** t in $\mathcal{T}$ **do** /* predict new locations of tracks */
    t ← KalmanFilter(t)
  **End**
  Associate $\mathcal{T}$ and $\mathcal{D}_{high}$ using Similarity#1 /* first association */
  $\mathcal{D}_{remain} \leftarrow$ remaining object boxes from $\mathcal{D}_{high}$
  $\mathcal{T}_{remain} \leftarrow$ remaining tracks from $\mathcal{T}$
  Associate $\mathcal{T}_{remain}$ and $\mathcal{D}_{low}$ using similarity#2 /* second association */
  $\mathcal{T}_{re-remain} \leftarrow$ remaining tracks from $\mathcal{T}_{remain}$
  $\mathcal{T} \leftarrow \mathcal{T} \setminus \mathcal{T}_{re-remain}$ /* delete unmatched tracks */
  **for** d in $\mathcal{D}_{remain}$ **do** /* initialize new tracks */
    $\mathcal{T} \leftarrow \mathcal{T} \cup \{d\}$
  **End**
**End**
Return: $\mathcal{T}$

# Vehicle Trajectory Extraction via Real-time Object Detection and Tracking in Video Analytics-Based Surveillance Systems

***Milad Bayat**, M.Sc., Grad., Department of Civil Engineering,*

*Iran University of Science and Technology, Tehran, Iran.*

***E-mail: MiladBayatAcademic@gmail.com***

Received: May 2026- Accepted: November 2026

**ABSTRACT**

A portion of the country's road network is equipped with surveillance cameras. These cameras are continuously monitored by human operators. Given the large number of cameras, this approach to traffic monitoring is not only labor-intensive and costly but also highly error-prone. Therefore, an automated framework with real-time processing capabilities is needed to extract vehicle trajectories that transportation researchers and practitioners can use to investigate microscopic vehicle behavior in various contexts. This study aims to design an advanced and efficient framework for extracting accurate and continuous vehicle trajectories using real-time object detection and tracking in video-analytics-based surveillance systems. In this study, the YOLOv10 algorithm was employed for object detection, while ByteTrack was used for object tracking. Recognizing the importance of access to a diverse, high-quality, and representative dataset, a dataset named Surveillance Cameras was developed from a variety of traffic scenes with the aim of providing high-quality annotations. Transfer learning was employed to fine-tune the pre-trained YOLOv10m model on this dataset. The fine-tuned model achieved precision, recall, and mean average precision values of 96.3%, 95.2%, and 98.3%, respectively, with an inference latency of 5.08 ms. The results demonstrate that the proposed framework can accurately detect and correctly classify most vehicles while ensuring a high frame rate in real-time applications. Within the tracking-by-detection paradigm, this performance improves the data association process and, while maintaining track-ID continuity across consecutive frames, enables the extraction of accurate, stable, and continuous vehicle trajectories. The proposed framework provides precise information on the positional changes associated with each track ID over time and can be applied to motion-pattern analysis, position prediction, traffic parameter extraction, and the identification of potential traffic conflicts.

**Keywords:** Computer Vision, Deep Learning, Real-Time Processing, Tracking-By-Detection, Trajectory Extraction